Using Channel-Specific Models to Detect and Mitigate Reverberation in Cochlear Implants

by

Jill M. Desmond

Department of Electrical and Computer Engineering
Duke University

Date: _____

Approved:

_____
Leslie M. Collins, Supervisor

_____
Lianne A. Cartee

_____
Lisa G. Huettel

_____
Jeffrey L. Krolik

_____
Loren W. Nolte

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Electrical and Computer Engineering in the Graduate School of Duke University
2014

Abstract

Using Channel-Specific Models to Detect and Mitigate Reverberation in Cochlear Implants

by

Jill M. Desmond

Department of Electrical and Computer Engineering
Duke University

Date: _____

Approved:

_____
Leslie M. Collins, Supervisor

_____
Lianne A. Cartee

_____
Lisa G. Huettel

_____
Jeffrey L. Krolik

_____
Loren W. Nolte

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Electrical and Computer Engineering in the Graduate School of Duke University
2014

Abstract

Cochlear implants (CIs) are devices that restore some level of hearing to deaf individuals. Because of their design and the impaired nature of the deafened auditory system, CIs provide listeners with limited spectral and temporal information, resulting in speech recognition that degrades more rapidly for CI listeners than for normal hearing listeners in noisy and reverberant environments [1]. This research project aimed to mitigate the effects of reverberation by directly manipulating the CI pulse train. A reverberation detection algorithm was initially developed to control processing when switching between the mitigation algorithm and a standard signal processing algorithm used when no mitigation is needed. Next, the benefit of removing two separate effects of reverberation was studied. Finally, two reverberation mitigation algorithms were developed. Because the two algorithms resulted in comparable performance, the effect of one algorithm on speech recognition was assessed in normal hearing (NH) and CI listeners.

Reverberation detection, which has not been thoroughly investigated in the CI literature, would provide a method to control the initiation of a reverberation mitigation algorithm. Although a mitigation algorithm would ideally remove reverberation without affecting non-reverberant signals, most noise and reverberation mitigation algorithms make errors and should only be applied when necessary. Therefore, a reverberation detection algorithm was designed to control the reverberation mitigation algorithm and thereby reduce unnecessary processing. The detection algorithm

was implemented by first developing features from the frequency-time matrices that result from the standard CI speech processing algorithm. Next, using these features, a maximum a posteriori classifier was shown to successfully discriminate speech in quiet, reverberation, speech shaped noise, and white Gaussian noise with 94% accuracy.

In order to develop the mitigation algorithm that would be controlled by the reverberation detection algorithm, a unique approach to reverberation mitigation was considered. This research project hypothesized that focusing mitigation on one effect of reverberation, either self-masking (masking within an individual phoneme) or overlap-masking (masking of one phoneme by a preceding phoneme) [2], may allow for a reverberation mitigation strategy that operates in real-time. In order to determine the feasibility of this approach, the benefit of mitigating the two effects of reverberation was assessed by comparing speech recognition scores for speech in reverberation to reverberant speech after ideal self-masking mitigation and to reverberant speech after ideal overlap-masking mitigation. Testing was completed with normal hearing listeners via an acoustic model as well as with CI listeners using their devices. Mitigating either effect was found to improve CI speech recognition in reverberant environments. These results suggested that a new, causal approach could be taken to reverberation mitigation.

Based on the success of the feasibility study, two initial overlap-masking mitigation algorithms were implemented and applied once reverberation was detected in speech stimuli. One algorithm processed a pulse train signal after CI speech processing, while the second algorithm processed the acoustic signal. Performance of the two overlap-masking mitigation algorithms was evaluated in simulation by comparing pulses that were determined to be overlap-masking with the known truth. Using the features explored in this work, performance was comparable between the two methods. Therefore, only the post-CI speech processing reverberation mitigation

algorithm was implemented in a CI speech processing strategy.

An initial experiment was conducted, using NH listeners and an acoustic model designed to present the frequency and temporal information that would be available to a CI listener. Listeners were presented with speech stimuli in the presence of both mitigated and unmitigated simulated reverberant conditions, and speech recognition was found to improve after reverberation mitigation. A subsequent experiment, also using NH listeners and an acoustic model, explored the effects of recorded room impulse responses (RIRs) and added noise (speech shaped noise and multi-talker babble) on the mitigation strategy. Because reverberation mitigation did not consistently improve speech recognition in these conditions, an analysis of the fundamental differences between simulated and recorded RIRs was conducted. Finally, CI listeners were presented with simulated reverberant speech, both with and without reverberation mitigation, and the effect of the mitigation strategy on speech recognition was studied. Because the reverberation mitigation strategy did not consistently improve speech recognition, future work is required to analyze the effects of algorithm-specific parameters for CI listeners.

Contents

List of Tables

List of Figures

Acknowledgements

Thinking back to the day I submitted my application to Duke University's Master's Program, I never imagined that I would leave here with a Ph.D. This turn of events is why my first "thank you" belongs to my advisor, Leslie Collins. Leslie's advisor role began before my acceptance to Duke, with her subtle suggestion that I change my application status to Ph.D. so that we could explore my options. Five years later, I think we thoroughly explored! Leslie, my life would not have been the same without your intervention, and I cannot thank you enough for encouraging me both initially and continually throughout my graduate school career.

Next up is another key advisor, Sandy Throckmorton, who helped me develop both the skills and the confidence required of a researcher. Sandy, whether you were reassuring me after one of my panicked late-night, pre-experiment emails or helping me brainstorm new directions when all seemed hopeless, you taught me a unique way of problem-solving. On the other end of the research process, you have also taught me the art of creating a solid presentation, and I'm confident that your feedback will be echoing in my head for many presentations to come. ("Make your own figures!" "Describe the process with a flowchart!' "Make your legends larger!")

I have also been lucky to work with some amazing labmates. First, the girl who showed me the CI-ropes: Sara Duran. Sara, you were not only an excellent mentor, holding my hand through my first CI experiments, but you have also been a great friend who was always there with a listening ear and encouraging words. Alterna-

tively, whenever I needed a good laugh, I could count on Ken Colwell and his practical jokes, like the infamous "USB-device-that-took-control-of-my-keyboard." Although that led to a very frustrating few months, I'm now ready to take on whatever practical jokes live outside these walls. I consider myself blessed to have been a member of such a strong and supportive lab, working with amazing engineers like Pete Torrione, Kenny Morton, Stacy Tantum, Mary Knox, Chris Ratto, Achut Manandhar, Rayn Sakaguchi, Jordan Malof, Patrick Wang, Boyla Mainsah, Kedar Prabhudesai, Dima Kalika, Nick Czarnek, Joe Camilo, Jillian Clements, and Daniel Reichman.

Without a doubt, this work would not have been possible without my research subjects: both those with and without implants. I am very lucky to have worked with such patient, committed, and enthusiastic CI listeners who, after hours of listening to tortuously noisy sentences, still left our sessions with a smile and with plans to return for another round. I would also like to thank the normal hearing listeners who signed up for my countless preliminary studies. Whether I was able to pay you in cookies or in actual dollars, I am very grateful for the hours you spent locked in the soundproof booth.

I wouldn't have made it this far in academia without a strong upbringing from two incredible parents, Sue and Marty Desmond. From childhood, I had a mother who knew every craft in the book (which resulted in the best Christmas gifts for friends and family). I grew up with a father who saw life through a child's eyes, inventing holiday traditions including running-pumpkins-over-with-the-family-station-wagon. I grew up with over-sized Christmas trees that only stood up after being tied to the window, days filled with bike rides around Southie (aka the "safest place in America"), and snowmen so big that only canoe paddles sufficed for arms. Most importantly, I grew up with parents whose support was endless. Whether I was deciding if I wanted to be an elementary school teacher or an engineer, or if I was deciding between grad school and starting a career, both of my parents have suffered

through countless hours of "well I don't knows" with a smile and a helping hand.

I am also very lucky to have grown up with a sister and a best friend, all wrapped into one amazing person. Leanne, whether I was following in your footsteps through high school Drama Club and the Marching Band, or whether I was flying across the world for a sister hiking trip to Mt. Everest Base Camp, I have cherished the many challenges that we have embarked on together, and I am excited to see what trouble the future holds. Between adventures, you have blessed me with a shoulder to cry on, someone to laugh with, and a person who understands me better than anyone else ever will. A sister-bond is one thing, but our "sistor" bond is like no other.

Last, but certainly not least, I would like to thank my companion, Justin Herman. Justin has gone above and beyond the call of boyfriend-duty: whether he was pouring me a glass of wine, listening to the woahs of my research for the thousandth time, or forcing me to take a much needed break when all I wanted to do was work through the night. Justin, you have not only provided me with endless support, but you have also been a source of endless entertainment. Whether you were busy filling in for a Hibachi chef and cooking dinner for friends and family, becoming a Waffologist for a local food truck, or surprising our visitors with a 15-passenger van to shuttle everyone to a local wine festival, you have made the past two years my favorite yet. Without you, graduate school would have lacked an essential amount of support, patience, and most of all, laughter.

1

Introduction

Reverberation, which is caused by the reflections of sound waves off of surfaces in the listening environment, results in delayed and attenuated reproductions of an original sound. Both normal hearing and hearing impaired listeners experience decreased speech intelligibility in reverberant conditions, resulting from smeared harmonic and temporal speech elements, blurred binaural cues, and flattened formant transitions [3; 4]. However, normal-hearing subjects often do not suffer decreased speech recognition until reverberation times (RTs) exceed approximately 1 second [e.g 3; 5], while an RT as low as 0.5 seconds may decease speech intelligibility for subjects with sensorineural hearing loss [e.g. 6; 7].

Although many cochlear implant users are able to function well in quiet conditions, receiving speech recognition scores of 80% or higher [8; 9], the addition of reverberation can be very detrimental to their speech intelligibility. Kokkinakis et al., 2011 found that a linear increase in RT results in an exponential decrease in speech intelligibility for CI listeners [7]. Therefore, this work aimed to mitigate the effects of reverberation in CI pulse trains.

Because reverberation decreases speech recognition for both normal hearing and

hearing impaired listeners, mitigating its effects has been studied in the acoustic literature. Unfortunately, many acoustic reverberation mitigation algorithms involve filtering the reverberant signal through the inverse of a room's impulse response (RIR) [e.g. 10; 11; 12; 13; 14; 15; 16]. These methods are not applicable for real-time CI processing, as their calculations are computationally demanding [e.g. 17]. Additionally, as the RIR depends on conditions such as room characteristics and source and microphone location, it must be updated frequently.

Reverberation mitigation strategies have also been studied in the implant literature. Kokkinakis et al., 2011 mitigated reverberation effects via a channel-selection strategy based on the signal-to-reverberant ratio within each frequency channel [7]. In a following study, Hazrati and Loizou, 2013 calculated channel-specific residual-to-reverberant ratios, using the residual signal resulting from the application of linear prediction strategies to the reverberant signal. An adaptive threshold was applied to these ratios to determine which channels should be retained and which should be discarded [18]. A third study, conducted by Hazrati et al., 2013, calculated channel-specific ratios of the variance of a speech signal raised to some power and the variance of the signal's absolute value. These features were then compared to an adaptive threshold to determine whether the given sample was dominated by speech or reverberation [19].

Although the aforementioned studies were shown to improve speech recognition in reverberation, real-time implementation of the strategies is not feasible. In the study presented by Kokkinakis et al., 2011, knowledge of the anechoic signal is required [7]. Alternatively, the work conducted by Hazrati and Loizou, 2013 requires condition-specific parameter tuning [18]. Finally, the algorithm presented by Hazrati et al., 2013 must contain knowledge of the future signal in order to update the adaptive threshold [19].

The goal of the current research project was to develop a causal reverberation

mitigation strategy that focused on statistical models of reverberant speech, rather than focusing on ratios of quiet and reverberant speech as attempted in previous studies. This project hypothesized that focusing mitigation on either the early or late reverberation reflections may result in a reverberation mitigation strategy that does not require RIR estimation and that is therefore feasible for real-time implementation.

Before a reverberation mitigation strategy was implemented, however, a reverberation detector was developed that operated on the simplified CI pulse train. This detector was developed to initiate a reverberation mitigation strategy, such that the strategy will minimally interfere with quiet CI processing [20]. Next, because this work aimed to mitigate either the early or late reverberation reflections, an initial study was conducted to investigate the possible impact of mitigating either effect. Specifically, speech recognition of unmitigated reverberant speech was compared to that in reverberation after ideal early- and late- reflection mitigation [21]. Although a similar study was completed by Kokkinakis and Loizou, 2011, their study only approximated the mitigation of each effect via manipulations of the vowel and consonant information. Specifically, to isolate self-masking and overlap-masking effects, respectively, reverberant consonants were replaced with clean consonants and reverberant vowels were replaced with clean vowels. [1]. Therefore, the ideal mitigation implemented in this work advanced the work of Kokkinakis and Loizou, 2011 by achieving results that are not influenced by the impact of vowel and consonant information on speech recognition.

Because it was unknown whether the additional information present in an acoustic signal would benefit reverberation mitigation, two reverberation mitigation strategies were developed: one operating on the acoustic signal and one that utilized the CI pulse train. Because comparable performance resulted from the two strategies, the following research project assessed the effect of the latter strategy on speech

recognition for both normal hearing (NH) listeners using an acoustic model and for CI listeners using their own devices. The aim of this research is to improve upon previous research by being both robust to noise, via the design of machine learning algorithms, and by operating in real-time, using signal features that depend only on current and previous time data.

2

Background

2.1 Cochlear Implants

In a normally hearing ear, small bones located in the middle ear convert pressure waves received by the outer ear into mechanical vibrations. As a result, fluid inside the cochlea (located in the inner ear) vibrates and causes the basilar membrane to shift. Hair cells that are attached to the basilar membrane bend in response to this shift. As a direct result of being bent, neurotransmitters are released from the hair cells, causing neighboring neurons to fire. A common form of deafness is caused by a loss of hair cells [e.g. 22]. However, if the nerves are intact, they can be excited with electrical stimulation and caused to fire.

The cochlear implant (CI) is a device designed to address the loss of hair cells. An array of up to 22 electrodes is inserted into the cochlea, and these are used to electrically stimulate the surviving neurons. In a normal hearing ear, the basilar membrane vibrates maximally at different locations given different stimulation frequencies, from low frequencies at one end (the apex) to high frequencies at the opposite end (the base) of the cochlea. Electrodes within the array of a cochlear implant are able to

make use of this tonotopic arrangement. Specifically, each electrode is responsible for transmitting the information corresponding to one frequency band (for a review, see [23]).

All cochlear implants are equipped with a microphone that sends sound to a speech processor (see Figure 2.1). The speech processor filters the input into the available frequency bands and converts the sound into electrical signals that will be transmitted to the implanted device either transcutaneously via a radio frequency (RF) signal or percutaneously (not shown). From here, the signal is decoded, transformed from a bit stream into current levels, and transmitted to the electrode array inside the cochlea. These electrodes, in turn, stimulate the auditory nerve, and this results in the perception of sound.

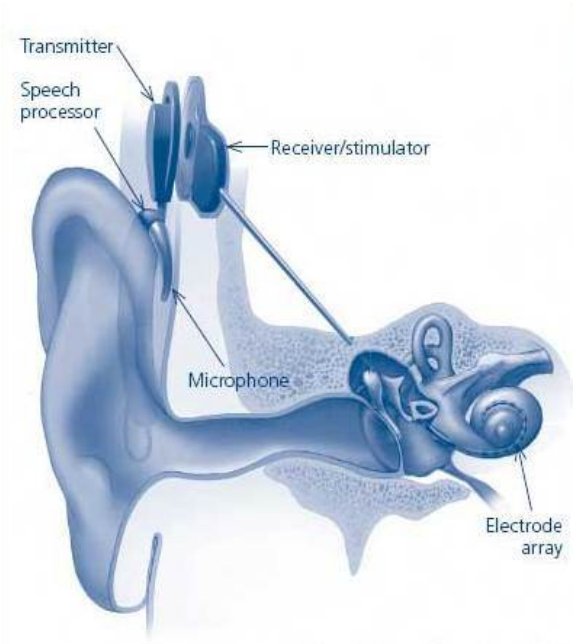

FIGURE 2.1: Diagram of a human ear with a cochlear implant. Sound is transmitted from a microphone to the speech processor. The signal, which is filtered into the available frequency bands and converted into electrical signals, is transmitted transcutaneously or percutaneously (not shown). The signal is then transformed into currents and sent to the electrode array located inside the cochlea, which is responsible for stimulating the auditory nerve (National Institutes of Health, Division of Medical Arts).

Although several speech processing algorithms are currently utilized in the cochlear implant population, the algorithm used in this work is the Advanced Combination Encoder (ACE) strategy. In this strategy (see Figure 2.2), sound travels from the microphone to an array of M bandpass filters, each corresponding to one electrode. The signal segments are then lowpass filtered and rectified in order to extract their envelopes. Next, the electrodes corresponding to the 'N' (less than 'M') frequency band envelopes with the greatest energy in each temporal analysis window are selected for stimulation. Following this step is an amplitude compression stage, which accounts for the reduced dynamic range of electric hearing. This signal then modulates a biphasic current pulse train, which is presented to the electrodes [e.g. 24].

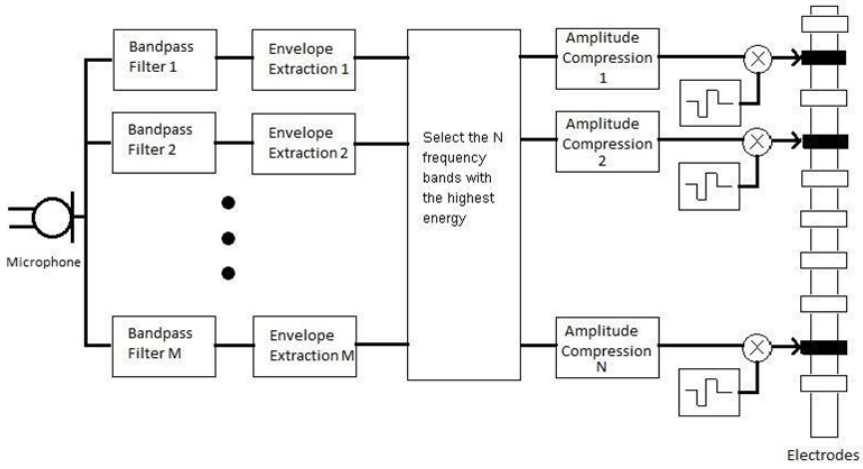

FIGURE 2.2: Diagram outlining the ACE processing strategy [e.g. 24]. Sound travels from the microphone to an array of bandpass filters. The envelopes are then extracted from the signal segments. Next, the electrodes corresponding to the 'N' frequency band envelopes with the greatest energy in each window are selected for stimulation. Following this step is an amplitude compression stage. Finally, the signal is modulated with a biphasic current pulse train, which is sent to the electrodes.

The speech processing algorithms determine the temporal and frequency information that will be presented to the cochlear implant user. This information can be visualized in plots termed electrodograms, as shown in Figure 2.3. Electrodograms are plots of amplitude at a given electrode location as a function of time. If an electrode is to be stimulated at a given time, a "tick" will appear with amplitude corresponding to the stimulation current level. Because each electrode corresponds to a different frequency band, electrodograms are a method of displaying the frequency and temporal content of the stimulation pattern, similar to a spectrogram for acoustic signals. High frequencies, which stimulate the base of the cochlea, are represented by lower numbers in the electrode array.

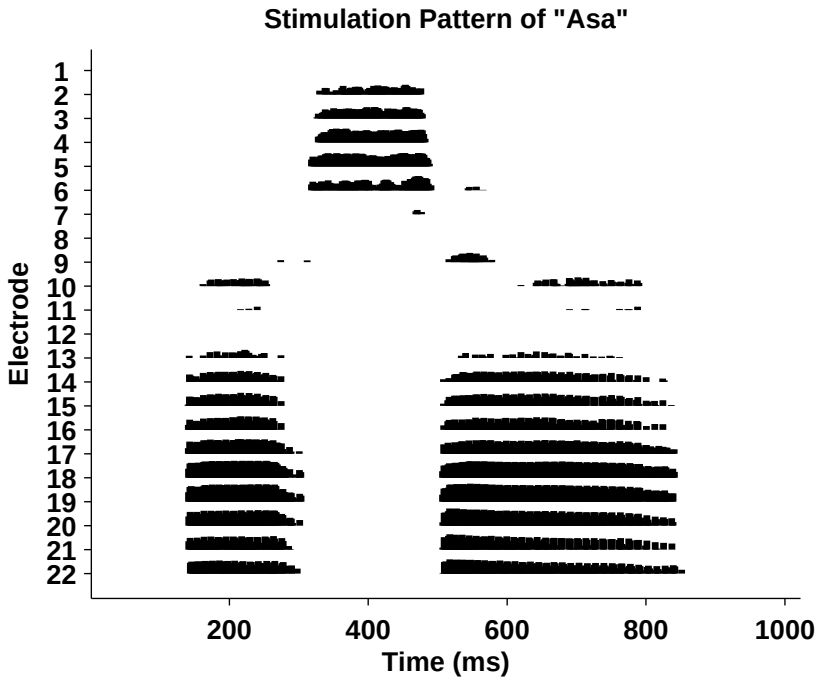

FIGURE 2.3: Cochlear implant stimulation pattern of the speech token "asa." This stimulation pattern, referred to as an electrodogram, demonstrates the frequency and temporal information that is presented to a cochlear implant user during the speech token "asa." Time is plotted on the x axis, and the y axis designates electrode number. If an electrode is to be stimulated at a given time, a "tick" mark, with amplitude corresponding to the stimulation current level, will appear at the corresponding location.

2.2 Reverberation

As previously mentioned, reverberation is especially detrimental to CI speech recognition. Reverberant speech, as observed by a microphone, can be described using Equation 2.1, where s(t), h(t), and n(t) represent the speech signal, the transfer function from the signal to the ear (the room impulse response, or the RIR), and the background noise, respectively, and * denotes convolution [e.g. 25]. The RIR is sensitive to many factors such as room dynamics (e.g. room size, shape, and surface materials), as well as the position of the listener and the sound source.

$$y(t) = h(t) * s(t) + n(t) \tag{2.1}$$

Another reverberation parameter, the reverberation time (RT), can be used to quantify the amount of reverberation present in a given room. The RT is defined as the amount of time required for a given frequency to decay to 60 decibels (dB) (relative to its original intensity) after the original sound is terminated. Reverberation time can be estimated using Equation 2.2, where V is the room volume (measured in ft^3) and $\Sigma S\alpha$ represents the sum of the surface areas of the materials in a given room (S, measured in ft^2), multiplied by their absorption coefficients at a given frequency (α) [26]. Most materials are poor at absorbing low frequencies, and as a result lower frequencies experience longer reverberation times [e.g. 27].

$$RT = \frac{0.049V}{\Sigma S\alpha} \tag{2.2}$$

Reverberation results in two main effects: self-masking (early reflections) and overlap-masking (late reflections) [4; 2]. Self-masking, which occurs during the first 50 ms following the source signal, alters the temporal and frequency information within an individual phoneme. Specifically, self-masking flattens formant transitions and flattens both the F1 and F2 formants, which can result in diphthongs being confused with monophthongs [4; 28]. This is especially detrimental to cochlear implant listeners, as they often find it difficult to perceive F1 and F2 formants in non-reverberant conditions. Kokkinakis and Loizou, 2011 hypothesize that the flattened formant transitions resulting from self-masking are the primary cause for speech intelligibility degradation for cochlear implant users [1].

Overlap-masking, on the other hand, results from reflections occurring greater than 50 ms following a source signal. This type of masking causes temporal smearing and can result in the reverberant sound from one phoneme masking a following

phoneme. Because vowels contain greater energy than consonants, overlap-masking has the potential to cause consonants to be masked by reverberant vowel data. In extreme conditions, entire words or parts of sentences may overlap, resulting in difficulty distinguishing the boundaries between words or sentences.

As mentioned previously, reverberation can hinder speech intelligibility for both normal hearing and hearing impaired listeners by smearing harmonic and temporal elements of speech, flattening formant transitions, and blurring binaural cues [4; 29]. To demonstrate some of these effects, Figure 2.4 displays an acoustic presentation of the sentence "She had your dark suit in greasy wash water all year", a sentence contained in the TIMIT database [30]. The top portion of this figure displays the acoustic waveform of the sentence in quiet, while the bottom signal includes reverberation with an RT of 1.2 seconds. This figure clearly illustrates the smearing of the temporal envelope that results from overlap-masking and that may disrupt phoneme and word boundaries.

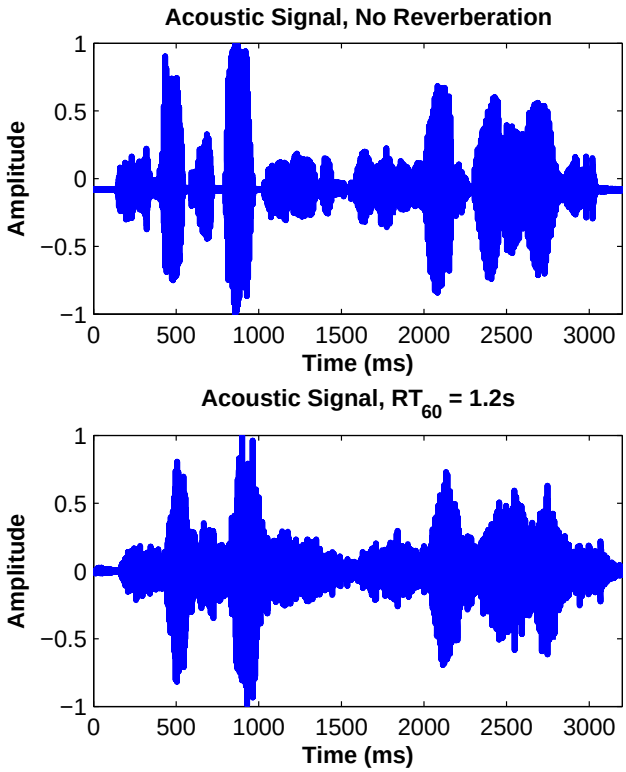

FIGURE 2.4: The sentence "She had your dark suit in greasy wash water all year," from the TIMIT database, in quiet (top) and in reverberation (bottom) with a reverberation time of 1.2s. This figure illustrates the smearing of word and phoneme boundaries that results from overlap-masking, as demonstrated in the bottom plot.

To visualize the effects of reverberation on a CI pulse train, Figure 2.5 displays electrodograms for the same sentence for which the acoustic waveform was plotted in Figure 2.4. This sentence was processed using the ACE coding strategy, in which nine electrodes were stimulated during each processing window. The top plot displays the electrodogram that results after the quiet signal was processed using the ACE strategy, while the electrodogram in the bottom plot was produced using the ACE strategy after reverberation was added with an RT of 1.2 seconds. The bottom plot of Figure 2.5 clearly displays smearing of the vowel-consonant boundaries. Although self-masking is also present, its effects are less easy to visualize in an electrodogram.

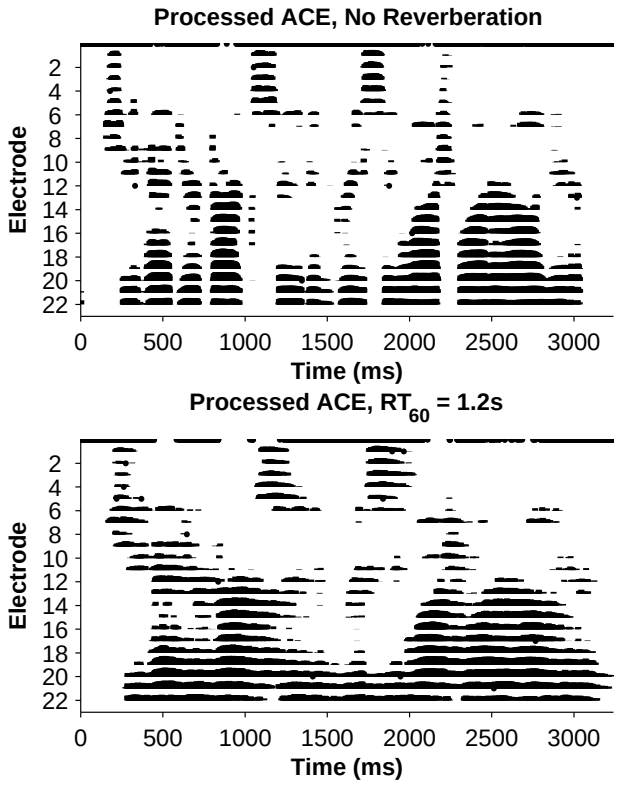

FIGURE 2.5: Cochlear implant stimulation pattern for the sentence "She had your dark suit in greasy wash water all year," from the TIMIT database, in quiet (top) and in reverberation (bottom) with a reverberation time of 1.2s. This sentence was processed using the ACE strategy, in which nine electrode channels were stimulated per window. The bottom figure demonstrates temporal smearing, which results in the loss of vowel-consonant boundaries.

2.2.1 Acoustic Reverberation Removal and Parameter Estimation

Because reverberation causes such detrimental effects to speech intelligibility, removing these effects in acoustic scenarios has been the focus of much research [e.g. 17; 31; 32]. One of the most common methods to reduce the impact of reverberation involves passing the reverberant signal through a filter designed to invert the effects of reverberation [e.g. 33; 34]. Unfortunately, these methods require estimations of the RIRs, which are difficult and computationally expensive to acquire [e.g. 17]. Additionally, the RIRs change frequently, as they depend on the room characteristics

as well as the position of the speaker and the listener.

Although challenging, many studies have attempted to estimate the RIR for reverberant acoustic signals. One study, conducted by Lin and Lee (2006), developed the Bayesian Regularization and Nonnegative Deconvolution (BRAND) algorithm to compute the RIR [35]. However, this algorithm assumes knowledge of speaker and microphone characteristics, information that may not be realistic in real-time computations. Another class of algorithms uses a test signal to predict the RIR [for a review, see 36]. Some test signal examples include the Maximum Length Sequence (MLS) [37], the Inverse Repeated Sequence (IRS) [38; 39; 40], Time-Stretched Pulses [41; 42], and a Sine Sweep, consisting of varied-frequency signals [43; 44]. Unfortunately, dependence on test signals is not feasible for real-time reverberation mitigation.

Other studies have focused on estimating the reverberation time of acoustic signals. Keshavarz et al., 2012 utilized linear predictive residuals and a maximum likelihood estimator to approximate the reverberation time [45]. Other methods involve a test signal, switching the signal off in order to measure the decay rate [e.g. 46]. The necessity of test signals in these algorithms limits their applications because, not only is it impractical to introduce test signals into real-world scenarios, but clean measurements of such test signals also cannot be guaranteed. Other methods that have been developed for determining RT are sensitive to the system's variables. One method, which is sensitive to speaker gender, utilizes a signal's periodicity to estimate RT [47]. Another algorithm, which is too complex for real-time implementation, estimates the power of a signal's envelope to predict reverberation time [48]. Yet another algorithm estimates the RT value by using a time-frequency decay model. However, both a long speech sample and speech immediately following a pause are required for the success of this method [49], and these speech tokens are not guaranteed in real-world listening environments. As a final example, Ratnam

et al., 2003 developed an algorithm that estimates the reverberation time using a maximum likelihood estimate of reverberant decay. However, this algorithm is highly susceptible to background noise [50].

Although acoustic reverberation has been the focus of many research studies, the results will not suffice for reverberation mitigation in cochlear implants. Acoustic RIR estimation is often too computationally expensive for real-time processing, assumes knowledge of some room characteristics, or requires a test signal for implementation. Reverberation-time estimation is also difficult for real-time implementation, as many algorithms utilize a test signal, are sensitive to system parameters, or are too complex to be implemented in real-time. However, in addition to the research conducted with NH listeners, there is also some research on reverberation mitigation in cochlear implants.

2.2.2 Removing Reverberation in Cochlear Implants

Many cochlear implant speech processing algorithms select only the frequency channels that have the greatest energy within an analysis window for stimulation. Because vowels (and their subsequent reverberation effects) often contain more energy than the subsequent consonants, cochlear implant processing strategies often select the reverberant vowel channels, rejecting the consonant channels altogether [7]. With this in mind, Kokkinakis et al., 2011 implemented a new channel selection criterion, which aimed to improve speech recognition for cochlear implant subjects in reverberant environments. This method used the signal to reverberant ratio (SRR), a measure of the ratio of signal energy from the direct and early reflections to the signal energy from late reflections. Channels with an SRR that existed below a threshold were rejected, eliminating temporal envelope smearing. The authors found that the new strategy improved speech intelligibility in reverberant conditions with an RT value of 1 second [7]. However, as the aforementioned method requires knowledge of the non-

reverberant target envelopes, more work is required to reduce real-time reverberation effects in cochlear implants.

Another study, completed by Hazrati and Loizou, 2013, uses a reverberant signal's linear prediction (LP) residual to mitigate reverberation. This study is motivated by the fact that the LP residual of a reverberant signal approximates the convolution of an anechoic signal's LP residual with the RIR [51; 52; 32]. At various time-frequency (TF) bins, the residual-to-reverberant ratio (RRR) was computed as the logarithm of the ratio of the residual signal's energy to the reverberant signal's energy. This ratio was compared against an adaptive threshold consisting of the weighted average of previous RRR values within the same frequency bin. RRR values larger than the threshold were dropped, as small RRRs suggest the presence of strong formant peaks. Although the reverberation mitigation algorithm implemented by Hazrati and Loizou, 2013 showed significant improvements in speech recognition over unprocessed reverberant speech, future work is still required. For example, the parameters associated with the adaptive threshold calculation are not optimal for all reverberation conditions and must be tuned for different configurations. Additionally, this algorithm may result in a delay depending on the processing power of the given device. Finally, their mitigation strategy only works at high SNRs, $SNR > 20dB$, because LP calculations vary in the presence of additional noise [18].

A separate reverberation mitigation strategy, the binary blind reverberant mask (BRM), was developed by Hazrati et al., 2013. This algorithm is applied to reverberant signals that have been binned into TF segments. Within each bin, the algorithm calculates the logarithm of the ratio of the variance of the signal raised to some power, to the variance of the absolute value of the signal. The exponent used in the feature calculation was determined experimentally. This feature was selected because of its similarity to kurtosis, which is lower in reverberant speech compared to anechoic speech [53; 54]. An adaptive threshold containing feature data

from 10 previous and 2 future frames was then applied to this feature, and TF bins with features less than the threshold were considered to be dominated by reverberation and were removed from the stimuli. The BRM algorithm resulted in significant speech intelligibility improvements for CI subjects at RT = 0.6s and 0.8s, but no statistically significant improvements were seen at RT = 0.3s. The lack of significant improvements at RT = 0.3s may be due to the fact that the BRM mitigates overlap-masking effects and, at low reverberation times, self-masking is dominant. Although promising, this algorithm requires future knowledge of the speech signal in order to calculate the adaptive threshold, which makes real-time implementation difficult. Additionally, the BRM loses information in the low frequency electrodes, further complicating speech intelligibility [19].

2.2.3 Current Approach to Reverberation Mitigation

A large body of research has investigated mitigating reverberation in acoustic signals using either estimates of RIRs or estimates of RTs. To date, these methods are not applicable to the real-time processing needs of CI speech processing due to the need to continuously update estimates under changing conditions or to present test signals to characterize the reverberant space. Research into algorithms specifically for mitigating reverberation in CIs are also to date not applicable for real-time processing. These algorithms rely on estimates of the non-reverberant signal, thus requiring non-causal features.

As an alternative to these methods of mitigating reverberation, a framework was developed that relied on machine learning to both control the onset of reverberation mitigation, thereby minimizing potential errors in low reverberation environments, as well as to apply the mitigation. The use of causal features for both reverberation detection and mitigation allows for real-time implementation without knowledge of the quiet signal or future information.

The rest of the document is structured as follows. In Chapter 3, the design of the reverberation detection algorithm is described, and the algorithm's performance under quiet, varying noise, and varying reverberant conditions is presented. This algorithm is envisioned as a switch to determine under what conditions the reverberation mitigation algorithm would be applied. In Chapter 4, the feasibility of the mitigation approach is investigated through ideal mitigation of self- and overlap-masking to verify that mitigating these effects independent of each other has the potential to improve speech recognition. Chapter 5 builds on the success of the feasibility study by developing two mitigation algorithms. Both algorithms aim to mitigate overlap masking but differ in the signals from which the features are extracted, with one algorithm based on the acoustic signal and the other based on the CI pulse train. While the acoustic pulse train has a much finer resolution in time and frequency, an algorithm based on the CI pulse train is more readily incorporated directly into the speech processing algorithm. Because comparable performance resulted from the two algorithms, only one algorithm was implemented for testing in subjects. The results from testing this algorithm under several conditions are presented in Chapter 6.

3

Reverberation Detection

Ideally, a reverberation mitigation algorithm would affect only the reverberant speech, leaving quiet speech unaltered. Because such an algorithm is difficult if not impossible to achieve, the goal of this research was to develop a reverberation detection algorithm that can be used to initiate a reverberation mitigation algorithm. An accurate reverberation detection algorithm would result in a mitigation algorithm that can be tailored specifically to reverberant speech.

Speech stimuli were degraded under a variety of simulated noise and reverberant conditions in order to develop and test the reverberation detection algorithm. This research used the CI pulse train, a signal with a much lower time and frequency resolution than the corresponding acoustic signal, to simplify modeling the statistical characteristics of reverberant speech. The simplified models were hypothesized to be less sensitive to reverberation condition changes such as head location and room dynamics. Further, by relying on the CI pulse train, implementation of the algorithms could occur within the CI speech processing algorithm rather than having to be applied as a pre-processing step. Therefore, all stimuli were processed into their CI pulse train representation. Features were then extracted from the signals to describe

characteristics specific to each class, and classification strategies were trained using these features. Finally, performance was evaluated for varying room and stimulation conditions.

3.1 Model Setup

3.1.1 Reverberation Room Model

In order to simulate the reverberation effects on an acoustic signal, RIRs for predefined rooms must be determined. In this experiment, a room was defined by the source and receiver position vectors, the room dimensions, and the reverberation time. Once an RIR was generated for a predefined room, the reverberant audio data was created by convolving the RIR with the source signal, as outlined in Equation 2.1. To approximate the RIRs, this work used the Modified Image Source Method (Modified ISM) technique, created by Lehmann and Johansson, 2008, based on the original ISM technique created by Allen and Berkley, 1978 [10; 55]. Simulated RIRs allowed various combinations of reverberation times, room dimensions, and source and microphone locations to be considered.

Using the original ISM technique, the RIR was calculated from the source to the receiver. This model uses image sources, located on mirror rooms which extend infinitely in all dimensions. Each image source contributes to the final signal in the form of a delayed and attenuated version of the original (source) signal. The sum of the power from the image sources distributed around the receiver is calculated to determine the power of the transfer function from the source to the microphone [10; 55].

The Modified ISM technique was developed to improve on the performance of the original technique. The original ISM method represents the RIR using a histogram, in which the bins represent discrete time values during which impulses were presented to the receiver by all image sources. One drawback of this technique is that

the time values must be rounded to fit into discrete bins, resulting in inaccuracies. Additionally, the original method requires a high-pass filter to allow the histogram to resemble an acoustic transfer function. To address these drawbacks, the Modified ISM technique operates in the frequency domain. In doing so, this model is able to accommodate time delays other than those at integer multiples of the sampling frequency [55].

Another alteration made by the Modified ISM technique involves the reflection coefficients β, which are defined for each surface in the given room. The reflection coefficients are calculated from the surfaces' absorption coefficients, α, as described in Equation 3.1. Differing from the original ISM technique, the modified technique utilizes the negative definition of the reflection coefficients [55].

$$\beta = \pm\sqrt{1-\alpha} \tag{3.1}$$

3.1.2 Cochlear Implant Stimulation Model

Once the RIR was created for a given reverberation time, it was convolved with an acoustic signal, resulting in the reverberant signal. This reverberant signal was then processed using the ACE speech processing strategy as presented in Section 2.1. This resulting reverberant cochlear implant pulse train was then processed by the reverberation detectors, which will be discussed in Section 3.2.2.

3.2 Methods

Although the ultimate goal of this work was to detect reverberation in cochlear implant pulse trains, it is important to ensure that speech in reverberation is differentiable from both quiet speech as well as from speech in other noise conditions. To begin, this research project classified sentences from the TIMIT database, created by Texas Instruments (TI) and the Massachusetts Institute of Technology (MIT)

[30]. These sentences were classified as existing in quiet, containing white Gaussian noise (WGN), containing speech shaped noise (SSN, noise that contains frequency characteristics of a long-term speech signal), or containing reverberation. Using MATLAB®, 3-15 dB (discretized in 2 dB increments) of WGN and SSN were added to the signals, while reverberation was added with an RT between 0.4 and 1.6 seconds (discretized in 0.2 second increments). The parameters were drawn from uniform distributions. The SSN was created using a 78th order finite impulse response (FIR) filter [56], with coefficients derived from an SSN sample supplied by the House Ear Institute. The WGN was created by randomly generating samples from a normal distribution in MATLAB®. Different instances of SSN and WGN were added to each TIMIT sentence used. For the reverberation simulation, a MATLAB® implementation of the Modified ISM technique, provided by Lehmann and Johansson, 2008, was used to create RIRs [55]. These RIRs were then convolved with the TIMIT sentences, via multiplication in the frequency domain, to create reverberant signals.

3.2.1 Modeling Speech in Cochlear Implant Pulse Trains

Both the quiet and noisy speech tokens were processed according to the ACE processing strategy, resulting in signals containing as many as 22 frequency bins. However, the process was interrupted prior to maxima selection, such that the stimuli from all active channels were modeled. Additionally, the stimuli used for classification were extracted prior to subject-specific dynamic range scaling. The result is an algorithm that is not influenced by dynamic range or maxima selection.

To model the activity in the frequency channels under various noise conditions, the timing between pulses, or the inter-stimulus intervals (ISIs), and the stimulation-lengths, or the duration (in ms) over which each channel remained active (was "on"), were considered for each channel in the ACE-generated frequency-time matrices. The presence of noise in a channel results in increased activity and decreased ISIs, because

shorter ISIs describe channels that do not remain off for large amounts of time. The stimulation-length distributions, on the other hand, were hypothesized to be negatively correlated with the ISI distributions. Therefore, locations of increased activity also experience increased stimulation-length values. Assuming that the state of the stimuli ("on" or "off") can be modeled as a Bernoulli random variable, the probabilities of the ISIs and the stimulation lengths were modeled as geometric distributions.

Different noise and interference scenarios should result in different activation patterns, allowing the aforementioned features to describe the signal characteristics. Because SSN is concentrated in the frequency bins associated with speech, its presence increases the amount of activity in the lower frequency regions. Although WGN is equally distributed across all frequencies, high-frequency channels contain more activity after the addition of WGN. This occurs because the cochlear implant channels are spaced logarithmically, to mimic the frequency arrangement of the cochlea. Because the high-frequency channels have a greater bandwidth, they also contain more WGN activity. Finally, because reverberation contains the original speech signal plus delayed and attenuated versions of this signal, it results in activation trends similar to quiet speech. However, more activity exists in each channel, resulting from the additional versions of the original stimuli that are present.

Figure 3.1 demonstrates the activation differences, in the form of normalized histograms of the ISIs, for a high frequency channel (left column) and a low frequency channel (right column). The TIMIT database was used to demonstrate speech in quiet (top row), as well as speech in 0 dB of SSN (second row), 0 dB of WGN (third row), and reverberation with an RT of 1.2s (bottom row). Figure 3.1 shows that SSN increases activity (decreases ISIs, with values closer to zero) in the speech-related low frequency channels, WGN increases activity in the high frequency regions, and reverberant speech resembles quiet speech but contains slightly shorter ISIs overall, resulting from additional reverberant stimuli.

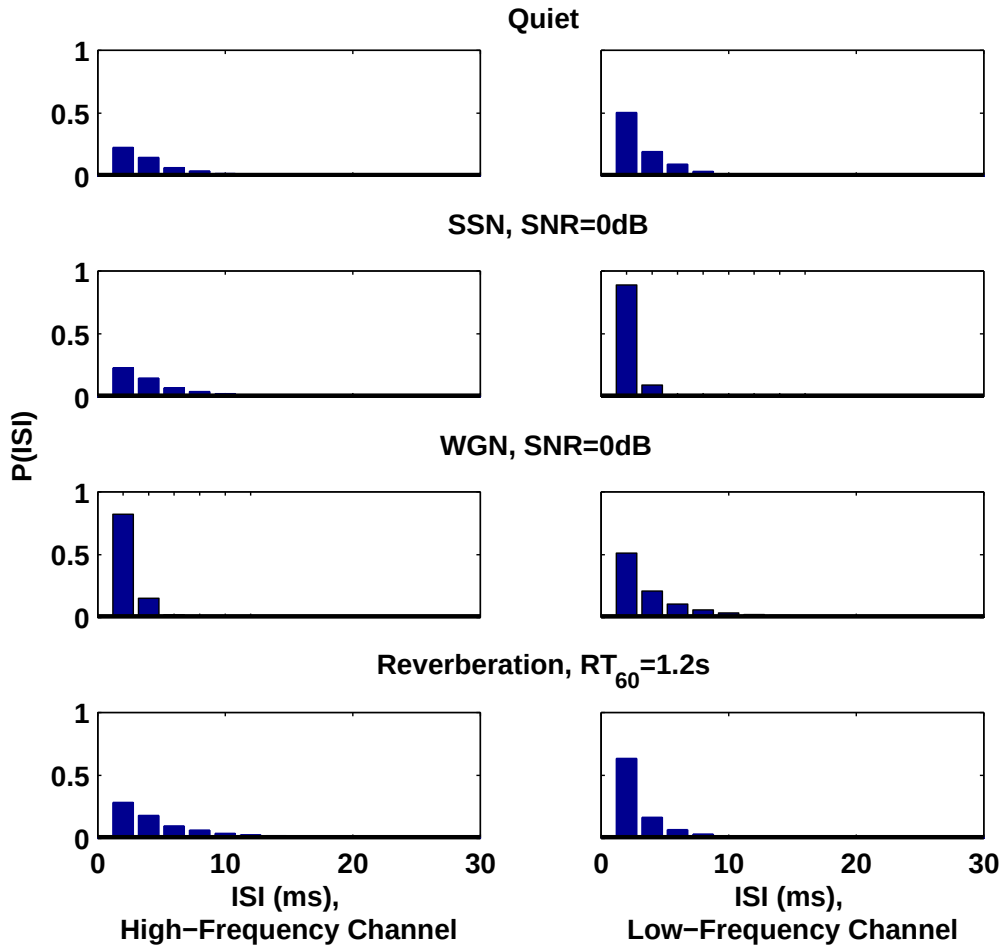

FIGURE 3.1: Normalized histograms describing ISIs for both a high frequency channel (left column) and a low frequency channel (right column) for speech in quiet (top row), SSN (second row), WGN (third row), and reverberation (bottom row). Shorter ISIs are apparent in the low frequency channels for quiet speech, speech in SSN, and speech in reverberation. Alternatively, shorter ISIs exist in the high frequency channels for speech in WGN, resulting from the logarithmic distribution of CI frequency bins [20].

The normalized histograms corresponding to a high frequency and a low frequency channel's stimulation-lengths are shown in Figure 3.2. (Note the scaling of the y-axis between 0 and 0.5). These features, which model the duration during which each channel remains "on" were expected to oppose the aforementioned ISI distributions, which describe the duration during which the channels were in an off state. An

example of this trend is visible for the low frequency channel in SSN (row 2, column 2): the ISI distribution is sharp, while the stimulation-length distribution experiences smearing. The high frequency channel in WGN (row 3, column 1) demonstrates the same effect. Although the two models are related, including both models improved the classifiers' performances, suggesting that both models contain some independent information.

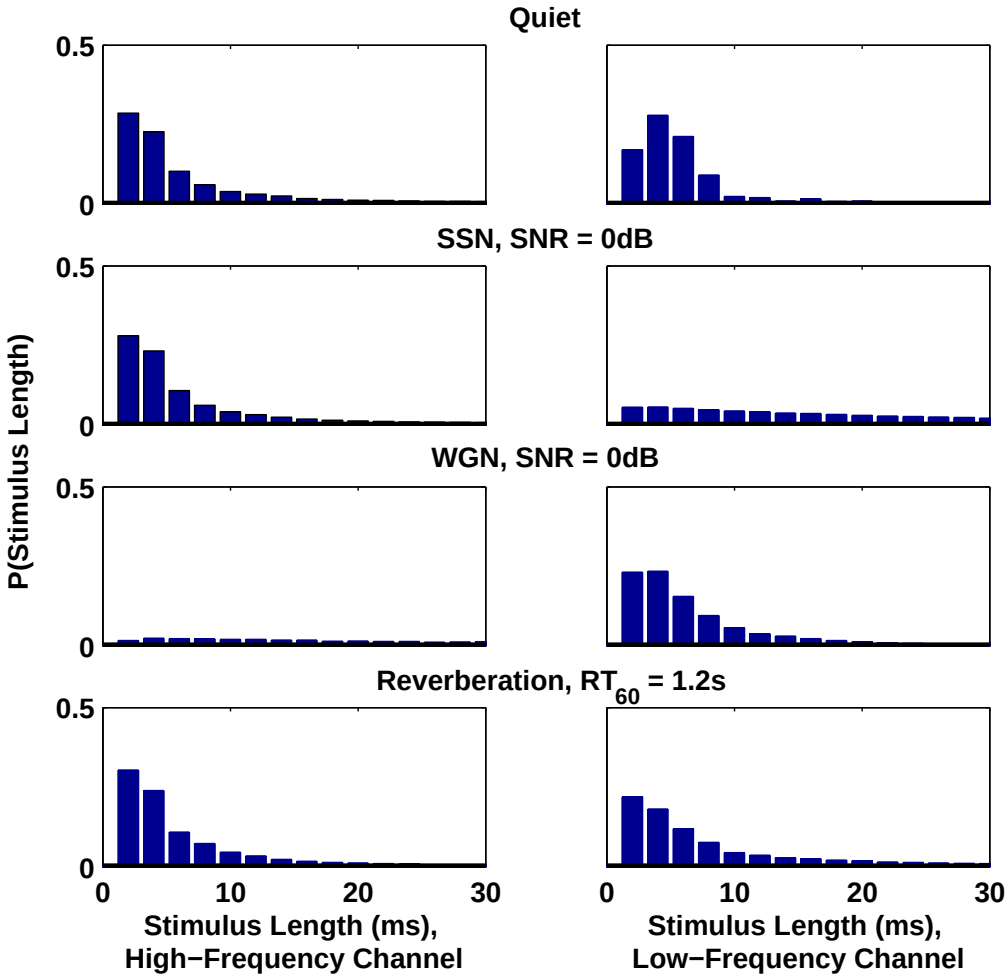

FIGURE 3.2: Normalized histograms describing stimulation-lengths for both a high frequency channel (left column) and a low frequency channel (right column) for speech in quiet (top row), SSN (second row), WGN (third row), and reverberation (bottom row). Smeared stimulation-lengths exist in the low frequency channels for quiet speech, speech in SSN, and speech in reverberation. However, as expected, longer stimulation-lengths exist in the high frequency channels for speech in WGN, resulting from to the logarithmic distribution of CI frequency bins [20].

3.2.2 Classification Algorithms

To describe the channel-specific models outlined in Section 3.2.1, geometric probability distributions were fit to each channel's ISI and stimulation-length data, and the resulting p-values were used as features to describe the speech models. Prior to

classification, features were normalized to have zero mean and unit variance.

Two classifiers, a maximum a posteriori (MAP) and a relevance vector machine (RVM), were considered. The MAP, a more generalized classifier, assumes that a multivariate normal distribution is adequate for describing the data. This assumption allows flexibility when presented with variable data, but the model could suffer if presented with data that does not follow the assumed distribution. The distributions resulting from the second classifier, the RVM, are formed from kernel functions placed at the feature locations. If the training and testing data do not vary significantly, the feature-specific distributions resulting from the RVM may be beneficial. Conversely, if the training and testing data do vary significantly, the RVM may result in over-training.

Cross-Validation

The classifiers considered in this research were trained and tested using ten-fold cross-validation. All TIMIT sentences were used for testing and training, and each TIMIT sentence was included in only one noise category (quiet, reverberation, SSN, or WGN). Cross-validation divides the available data into ten groups, or folds, with approximately equal representation for each noise condition in each fold. During each iteration, nine folds are used to train the classifiers, and the remaining fold is used for testing purposes. Each fold acts as the testing fold for one iteration, and this process is completed ten times [e.g 57].

Classification

First, a MAP classifier was used to detect reverberation. Given the features, this classifier selects the hypothesis that maximizes the posterior distribution. [e.g. 57]. Using maximum likelihood estimation to calculate the mean and covariance matrices, a multivariate normal distribution was assumed to describe each class' features.

Next, a kernel-based classifier, the RVM, was implemented as a classification algorithm. After placing kernel functions at the training point locations, the RVM creates sparsity by removing or "pruning" some of the less-informative kernel functions [58]. Gaussian radial basis functions containing a width of one were used as kernels, and DC kernels were also included to account for any offsets in the data.

3.3 General Reverberation Detection

3.3.1 Stimuli

Cochlear implant stimulation parameters are subject-specific. The classifiers were first tested using a general set of cochlear implant clinical parameters with a pulse rate of 800 pulses per second (pps) and 22 active electrodes.

The speech samples were created in quiet, in speech shaped noise (SSN), in white Gaussian noise (WGN), and in reverberation. Noise levels varied, discretized in 2 dB increments, between 3-15 dB for SSN and WGN, and reverberation was simulated with an RT varying between 0.4s and 1.6s. The RIRs were simulated with a room dimension of (10.0 x 6.6 x 3.0)m, a source location of (2.4835 x 2.0 x 1.8)m, and a microphone located at (6.5 x 3.8 x 1.8)m, as used by Champagne et al., 1996 [59]. The room dimension was selected such that it was large enough to contain adequate reverberation, but small enough to be applicable to everyday situations.

3.3.2 Results

The labels estimated by the classifiers and the known class labels were used to score the results for accuracy. The classification results generated by the MAP and RVM classifiers are provided in the confusion matrices in Figure 3.3. In a confusion matrix, correct classification categories are displayed across rows (top to bottom: speech in quiet, speech in SSN, speech in WGN, and reverberant speech) and classifier assignments are displayed down columns (left to right: speech in quiet, speech in

SSN, speech in WGN, and reverberant speech). Percentage values across the diagonal of the figures represent correct classifications, while the remaining squares represent incorrect classifications. As seen in the left plot of Figure 3.3, the MAP classified reverberation 91.7% of the time it was present, with an overall detection accuracy of 91.14% across all signal classes. The RVM (right), on the other hand, correctly identified reverberation 96.2% of the time it was present, with an overall accuracy of 91.48%.

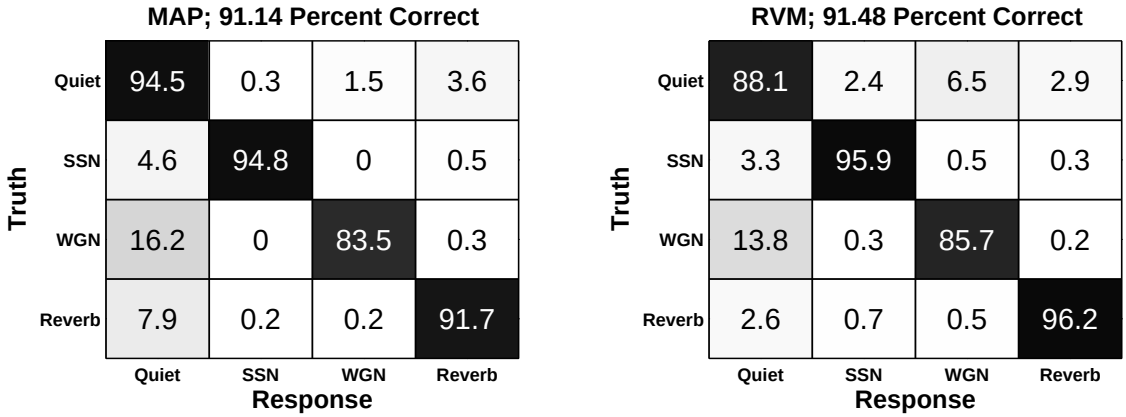

FIGURE 3.3: Confusion matrices displaying the MAP (left) and RVM (right) classification results for reverberant data created with RT varying between 0.4s and 1.6s, in intervals of 0.2s. The remaining reverberant room parameters were assigned as in Champagne et al., 1996 [59]. In the figures, rows display the correct classification labels (truth), while columns represent classifier assignments (response). The diagonals contain the percentages of correct classification, while incorrect classification percentages appear in the segments corresponding to each signal's actual and assigned labels. According to these confusion matrices, reverberation was accurately classified 91.7% and 96.2% of the time it was present for the MAP and RVM classifiers, respectively. The overall accuracy across all signal classes was 91.14% for the MAP and 91.48% for the RVM [20].

Application of both classifiers resulted in similar performance, and reverberation was not overly confused with the remaining noise categories. Using the aforementioned specific listening conditions, the ISI and stimulation-length features resulted in good discrimination. However, classification was completed assuming a generic set

of CI subject clinical parameters. Because, in reality, each CI listener has a unique set of parameters that could affect the performance of the reverberation detection algorithms, a sensitivity analysis was conducted to investigate the effect of these parameters on performance.

3.4 Classifier Robustness to Subject Clinical Program Parameters

3.4.1 Stimuli

Each cochlear implant listener has a unique set of parameters resulting in subject-specific stimulation pulse trains. Some parameters, for example the subjects' dynamic ranges and the number of channel maxima stimulated per time window, have no effect on the reverberation detection performance because the implant pulse trains were processed before applying these variables to generate the final stimulation patterns. Other parameters, which alter the signals presented to the classifiers, could affect performance. These parameters include the set of channels selected for stimulation, the channel stimulation rate, and the equation mapping current (in μA) to cochlear implant current steps, shown in Equations 3.2 and 3.3. Current steps, used by Cochlear Corporation to define the amount of current presented to the electrodes, are represented by CL in Equations 3.2 and 3.3.

$$I(\mu A) = 10e^{\frac{CL \cdot ln(175)}{255}} \tag{3.2}$$

$$I(\mu A) = 17.5 \cdot 100^{\frac{CL}{255}} \tag{3.3}$$

To test the algorithms' sensitivity to different clinical parameters, 100 simulated configurations were created with varying parameters. Between 18 and 22 channels were selected at random, the channel stimulation rate was randomly assigned a value between 500 and 1200 pps (discretized in 100 pps increments), and the current-

mapping equation was randomly determined. As a result of randomly generating parameters, duplicate parameter configurations may exist. Each set of parameters was then used to process all TIMIT database sentences, and the MAP and RVM reverberation detection algorithms were applied to the data using ten-fold cross-validation for each parameter set separately. Results were compared to the those presented in Figure 3.3, which utilized 22 channels, a stimulation rate of 800 pps, and Equation 3.3 to map the current in μA to current steps.

3.4.2 Results

Figure 3.4 displays histograms of the MAP and RVM classification performance across noise types, using the varied subject parameters described in Section 3.4. When using the set of subject clinical parameters described in Section 3.3.1, the MAP and the RVM correctly classified all signals with accuracies of 91.14% and 91.48%, respectively. Varying the subject stimulation parameters resulted in performance comparable to the performance observed for the original fixed subject stimulation parameters.

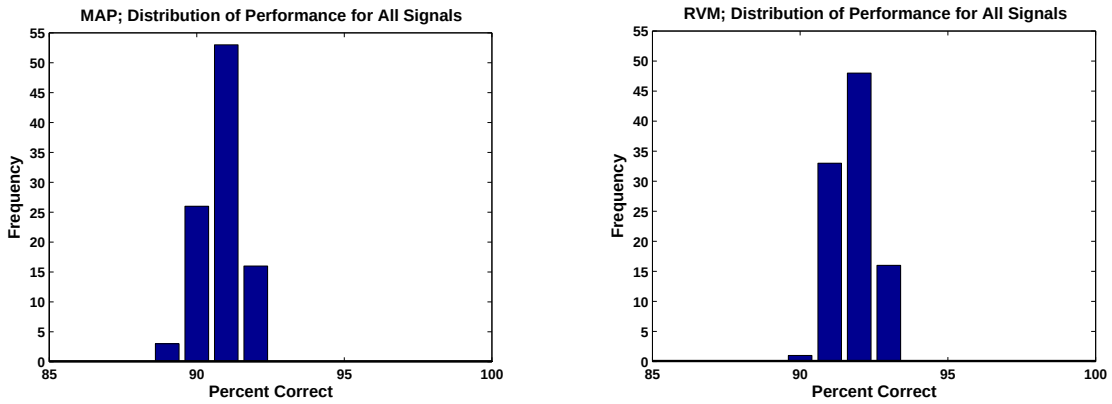

FIGURE 3.4: Histograms displaying the percentage of correctly labeled signals, across all noise types, for the MAP (left) and RVM (right) classifiers. These results were trained and tested using 100 random subject stimulation parameter configurations. (The training and testing was completed for each parameter configuration separately.) For each parameter set, the detection algorithms classified signals as existing in quiet, or containing SSN, WGN, or reverberation. These results are comparable to the results generated with the original subject parameters described in Section 3.3.1 [20].

Histograms displaying the percentage of reverberant signals correctly labeled by the MAP and the RVM are displayed in Figure 3.5. Results determined when using differing parameters are comparable to those using the general stimulation parameters, which had an accuracy of 91.7% for the MAP and 96.2% for the RVM. The MAP results vary more substantially than the RVM results, which could be due to the naive MAP classifier distributions, which assume that features can be described by a multivariate normal distribution, compared to the more precise distribution that results from the application of kernels in the RVM.

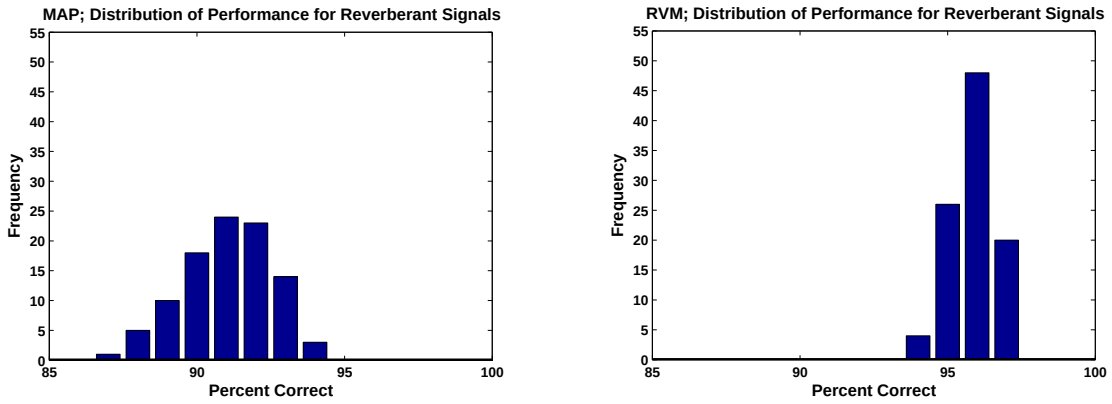

FIGURE 3.5: Histograms representing the MAP (left) and RVM (right) reverberation detection accuracy. 100 randomly generated subject parameters were used to generate the frequency-time matrices used for classification, and training and testing occurred on each parameter set separately. Performance does not appear to suffer compared to the results created using one set of subject stimulation parameters. [20]

The results found when varying subject parameters, seen in Figures 3.4 and 3.5 were comparable to those that resulted from the general parameters, suggesting that performance is robust to differing clinical parameters. Although seemingly robust to subject clinical parameters, many of the reverberation parameters in this and the previous section were fixed, assuming known room characteristics. In a more realistic listening scenario, the room dimensions and the locations of the source and microphone may be unknown. Therefore, performance when presented with different room parameters was investigated.

3.5 Classifier Robustness to Room Configurations

3.5.1 Experimental Setup

In this study, the classifiers' robustness to various room configurations was tested by randomly varying the parameters used to generate the RIRs. First, the impact of varying the room dimensions, the positions of the source and microphone, and the RT was investigated (see Section 3.5.2). This experiment was followed by experiments to test the sensitivity of the classification algorithms to specific reverberation

parameters. Scenarios were created in which each parameter was set to a constant, allowing the remaining parameters to vary randomly.

Because allowing the reverberation parameters to vary increases the level of difficulty for the classification algorithms, three sentences from the TIMIT database were concatenated for each training and testing speech token in the following sections. This ensures that enough data was present given the increased level of difficulty. Because the following results were generated with additional sentences in each training and testing feature set, the classifications cannot be compared directly with the results reported previously. In the remaining sections, using ten-fold cross-validation and concatenated sentences, the classifiers had access to 675 training and 75 testing groups in each fold. There was no overlap between the training and testing folds.

3.5.2 Results: Varying All Parameters

Room dimensions were generated between (2 x 2 x 2)m and (50 x 50 x 50)m (in an effort to create realistic room configurations, length and width were within a factor of 2 of each other, and the height could not be greater than twice the length or width). The source and the microphone were also randomly positioned within the room, and the reverberation time varied varied between 0.4s and 1.6s.

The confusion matrix that resulted from application of the MAP classifier to this data is shown in the left column of Figure 3.6, while the result of the RVM classifier is displayed in the right column. The MAP classifier correctly detected reverberation in 93.7% of the signals in which it existed, and the classifier's overall accuracy was 90.48%. The RVM resulted in worse performance, with 86.8% reverberation detection accuracy and 88.36% accuracy across signals. However, RVM parameters, such as the radial basis functions describing the kernel functions, were not optimized, and it is possible that the RVM performance may increase with post hoc optimization. The MAP classifier is advantageous in that the parameters do not require optimization.

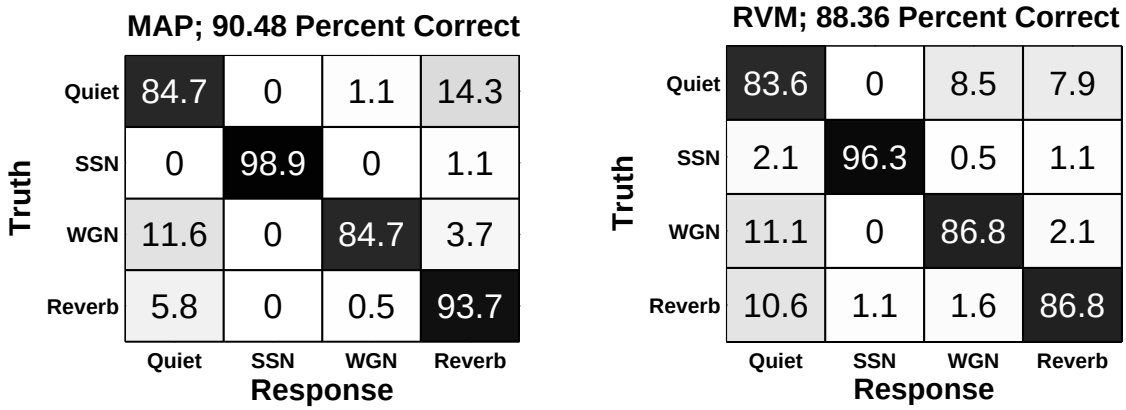

FIGURE 3.6: Classification performance of the MAP classifier (left) and the RVM classifier (right) in the presence of varying reverberant conditions, with room dimensions varied from (2 x 2 x 2)m to (50 x 50 x 50)m, source and microphone locations randomly positioned within the room, and RT varied from 0.4s to 1.6s. The MAP classifier correctly identified reverberation in 93.7% of signals and correctly labeled all signals with an accuracy of 90.48%. The RVM correctly classified reverberation in 86.8% of reverberant signals, with across-class accuracy of 88.36% [20].

3.5.3 Results: Impact of Each Parameter on Classification

To test the impact of the reverberation time, the MAP and RVM classifiers were applied to data in which the RT was fixed and the remaining parameters varied as in Section 3.5.2. RT was set to either a relatively low level of reverberation (0.5s) or a relatively high level of reverberation (1.2s). The top row of Figure 3.7 displays the classification performance resulting from an RT value of 0.5s. Unsurprisingly, the low RT impedes performance, with the MAP and RVM classifiers correctly detecting reverberation with 85.7% and 72.5% accuracy, respectability. Alternatively, increasing the RT to 1.2s increased performance for both the MAP and RVM classifiers, shown in the second row of Figure 3.7. The difference in detection performance between the low and high reverberation times suggests that this parameter greatly impacts the algorithms' performance.

The room dimensions were then fixed to (10.0 x 6.6 x 3.0)m, as described by Champagne et al., 1996 [59], and the remaining parameters varied as described in

Section 3.5.2. The results for these conditions can be seen in the third row of Figure 3.7. Because the performance improved over varying all room parameters (see Figure 3.6), knowledge of the room layout appears to improve reverberation detection performance. However, fixing the microphone position (row 4, Figure 3.7) or the source position (row 5, Figure 3.7) resulted in little accuracy increase when compared to varying all room parameters (Figure 3.6). The source and microphone positions are known to influence the room impulse response, but their impact appears to be reduced when considering the pulse train stimuli presented to a cochlear implant.

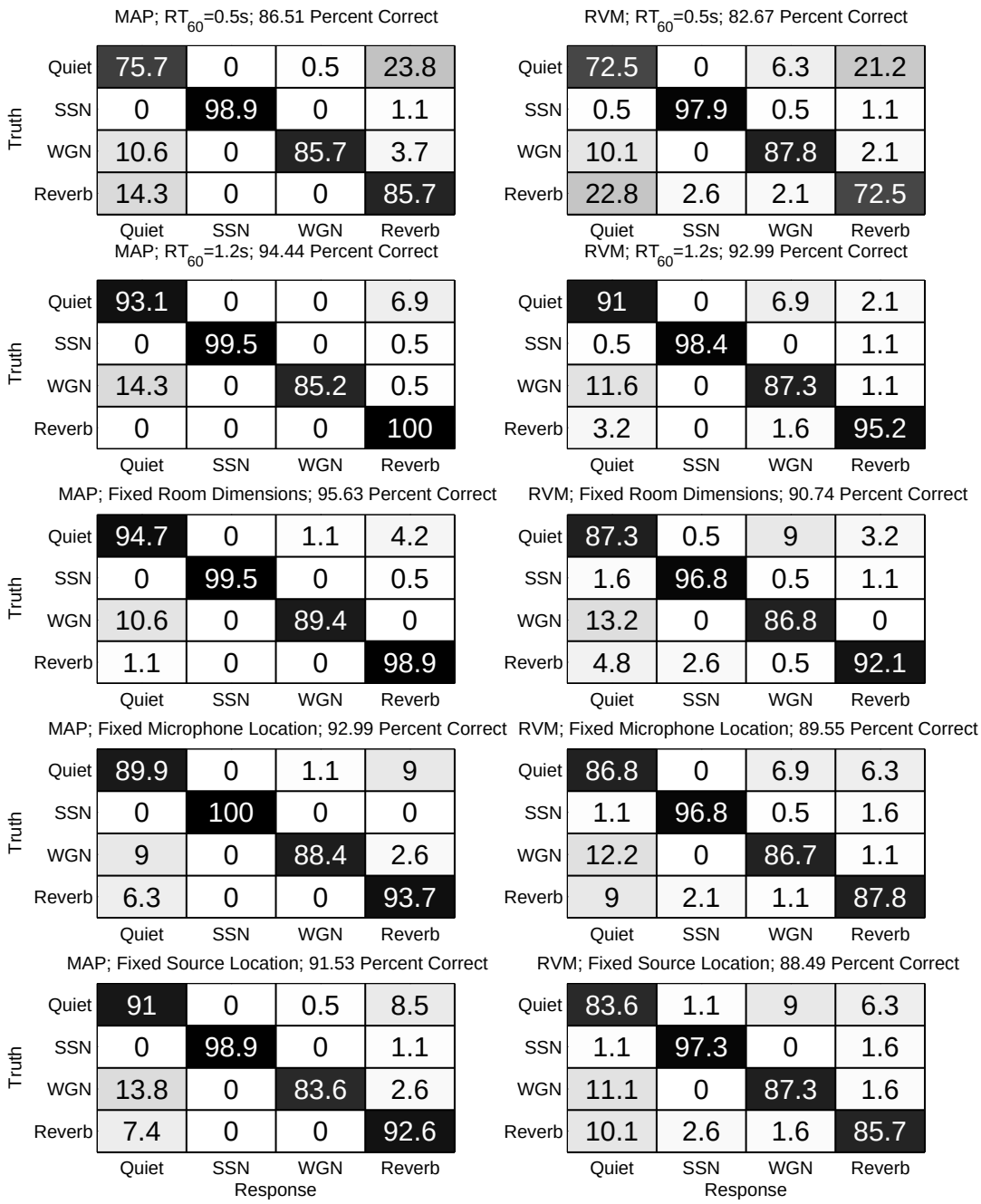

FIGURE 3.7: MAP (left column) and RVM (right column) classification results applied to data in which reverberant signals had a fixed RT of 0.5s (top row), a fixed RT of 1.2s (second row), room dimensions fixed at (10.0 x 6.6 x 3.0)m, as used by Champagne et al., 1996 (third row), microphone location positioned at the room's center (fourth row), or source location held at the center of the room (fifth row). The reverberation detectors seem to suffer at low reverberation times, and the locations of the source and microphone seem to have the smallest impact on the reverberation detection performance. [20]

3.6 Performance in Combined Reverberation and Noise

3.6.1 Experimental Setup

Because real-world listening environments often include speech samples containing both reverberation and noise, the performance of reverberation detection in the presence of either SSN or WGN was evaluated. Noise samples were added prior to the application of reverberation, and speech signals were classified in three groups: quiet speech, speech in the presence of noise, or speech in the presence of noise and reverberation. Three sentences were concatenated from the TIMIT database for each testing and training feature set. Noisy speech was created by adding WGN or SSN to quiet speech samples, with SNRs ranging from 3-15 dB (discretized in 2 dB increments). Noisy speech in the presence of reverberation was created by introducing either WGN or SSN (with SNRs of 3-15 dB, discretized by 2 dB increments) to the quiet tokens, followed by the addition of reverberation, with room characteristics randomized as described in Section 3.5.2.

3.6.2 Results

Performance of the classifiers applied to quiet speech, noisy speech, or noisy speech in a reverberant environment is shown in Figure 3.8. With varying parameters such as noise type, noise level, and reverberation condition, data within each class vary considerably. When presented with this challenging data, the MAP and RVM classifiers succeed in detecting reverberation in 87% and 86.5% of signals, respectively. Overall classification performance has 86.24% and 83.33% accuracy for the MAP and RVM classifiers, respectively.

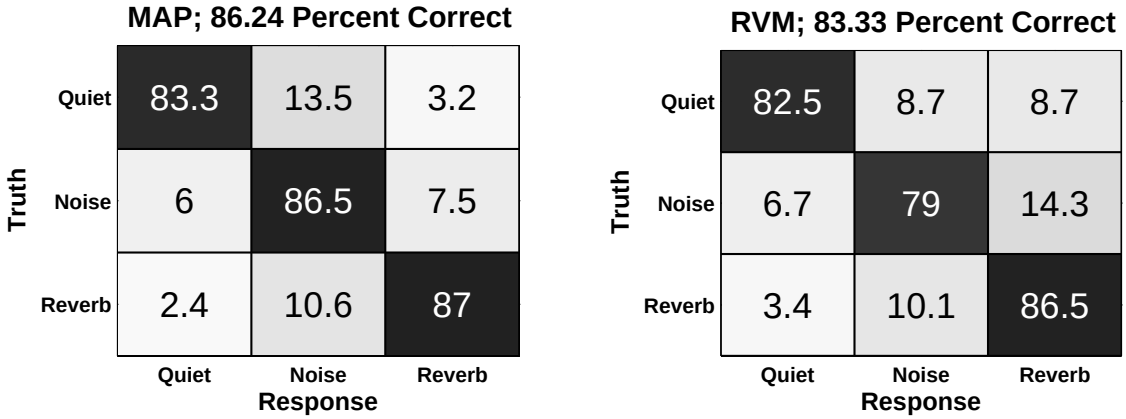

FIGURE 3.8: Confusion matrices displaying the classification performance of the MAP classifier (left) and the RVM classifier (right) when presented with varying noise and reverberation conditions. The categories for classification include quiet speech ("Quiet"), speech with the addition of either SSN or WGN ("Noise"), and speech with the addition of either SSN or WGN in the presence of reverberation ("Reverb"). The MAP correctly detected reverberation with 87% accuracy, and had an overall across-class detection accuracy of 86.24%. The application of the RVM resulted in 86.5% reverberation detection accuracy and 83.33% across-class classification accuracy [20].

3.7 Discussion

This research differs from previous research that often focused on estimating either an acoustic signal's reverberation time or room impulse response. These algorithms are not only computationally demanding, but they also assume the presence of reverberation and must be recalculated as room parameters change. Because the cochlear implant pulse train is lower in time and frequency resolution than the acoustic signal, the CI pulse train was used to detect reverberation with changing room characteristics, with the hypothesis that the CI pulse train might be less sensitive to changing reverberation parameters.

ISI and stimulation-length features were used as classifier inputs in order to discriminate reverberant speech from speech in quiet, speech in SSN, and speech in WGN. The features also seemed to describe reverberation in the presence of noise,

as reverberant speech in the presence of either SSN or WGN was detectable. Additional features, such as the amplitudes of pulses, the changes in consecutive pulse amplitudes, and the total pulse count in each channel were also considered. These features were not included in the final classification, as they did not consistently improve performance of the classifiers.

Higher values of RT increased performance, as did knowledge of the room dimensions. Prior information about the source and microphone position, on the other hand, had less of an effect on detection performance, suggesting that these parameters have little effect on the cochlear implant activation patterns.

The performance of the reverberation detection algorithms was highly dependent on reverberation time, with a change in RT from 1.2 seconds to 0.5 seconds resulting in a drop in MAP classifier detection performance from 100% to 85.7%. Much of this performance decrease was due to misclassification between reverberant speech and quiet speech, with quiet speech labeled as containing reverberation in about 14% of cases and reverberant speech labeled as quiet speech in about 20% of cases. Labeling quiet speech as reverberant speech will result in a mitigation algorithm incorrectly being activated. The impact of this misclassification will depend on the errors introduced in the quiet speech pattern by the reverberation mitigation algorithm and requires further investigation to determine whether the impact would be significant. On the other hand, incorrectly labeling reverberant speech as quiet will simply result in not initiating a reverberation mitigation algorithm when reverberation is present. Kokkinakis et. al., 2011 suggest that reverberation containing a reverberation time of 0.5 seconds can result in speech recognition scores decreasing from 90% correct to 50% correct. If a 14% miss rate of detecting reverberation is considered, a 6% drop in speech recognition might be hypothesized to result from failing to mitigate, assuming ideal mitigation when reverberation is detected. Although it is expected that reverberation detection performance will decrease as reverberation time drops

below 0.5 seconds, the impact should be minimized due to an increase in speech recognition performance of cochlear implant subjects in less reverberant conditions.

Because this research only considered noise and reverberation scenarios generated in simulations, future research would include testing the algorithms with recorded noise samples and room impulse responses. Because recorded room impulse responses contain more variable frequency responses, they may introduce more challenging listening situations. However, the results presented in this chapter suggest that reverberation is not only detectable in cochlear implant pulse trains, but may also be robust to changing room dynamics and cochlear implant stimulation parameters. Reverberation detection provides a key first step in the efforts to mitigate reverberation effects for CI listeners. Once reverberation has been detected in an implant pulse train, a reverberation mitigation algorithm can be initiated. In order to develop a real-time reverberation mitigation strategy, this research project aimed to mitigate either self- or overlap- masking. Before such a mitigation algorithm could be developed, however, the potential for speech intelligibility improvement after ideal self- or overlap- masking mitigation was investigated in a feasibility study, presented in Chapter 4.

4

Effects of Self- and Overlap- Masking on Speech Intelligibility

As mentioned in Section 2.2, reverberation results in self-masking (masking within an individual phoneme) and overlap-masking (the masking of one phoneme by a preceding phoneme). The ultimate goal of this research was to mitigate the effects of reverberation, and focusing on either self- or overlap- masking may provide a tangible first step. To determine whether mitigating either of these two effects independently of each other has the potential to improve speech recognition, a feasibility study was conducted using ideal mitigation.

4.1 Subjects

The use of human subjects in the experiments associated with this work was approved by The Institutional Review Board of Duke University, and the participants of this experiment were compensated for their time.

4.1.1 Normal Hearing Subjects

Four female and six male normal hearing subjects were recruited for an initial experiment. Reverberant speech was presented using an acoustic model that approximates the frequency and temporal information that is available to a CI listener [60]. The experiment required one test session lasting approximately 45 minutes.

4.1.2 Cochlear Implant Subjects

Four CI subjects, all postlingually deaf and users of Cochlear Corporation's implants, were recruited for this experiment. Their demographic information is presented in Table 4.1. The ACE processing strategy, which is used by all participants, was used for the speech recognition tasks in this experiment, and only electrodes that were active in each subject's clinical parameter set were used. This experiment was completed in one approximately 3 hour session.

Table 4.1: Demographic information for the implanted subjects.

Subject ID	Gender	Age (years)	Age at onset of deafness (years)	Age at implantation (years)	Implant type
S1	M	71	55	67	CI24RE
S2	M	59	48	49	CI24R
S3	M	61	15	53	CI24RE
S4	F	50	10	41	CI24RE

4.2 Experimental Design

4.2.1 RIR Generation

Similar to the methods used in Chapter 3, the Modified ISM technique was used to approximate the RIRs [55]. RIRs were generated with varying RTs, while other parameters remained fixed. The room dimensions were set to (10.0 x 6.6 x 3.0)m,

the source location was (2.4835 x 2.0 x 1.8)m, and the microphone was located at (6.5 x 3.8 x 1.8)m [59].

4.2.2 Isolating the Effects of Reverberation

In order to isolate the effects of reverberation, sentences were created that contained either solely self-masking or solely overlap-masking effects. To label sections of self- and overlap- masking, both the unmitigated reverberant speech and the original non-reverberant speech were used. The reverberant and non-reverberant stimuli were smoothed on a channel-by-channel basis post-CI processing, using a summing filter that spanned five windows. (In this experiment, CI window lengths were 2 ms in duration, resulting in a summing filter spanning 10 ms). For both stimuli, windows resulting in sums greater than each subjects' channel-specific thresholds were labeled as speech tokens, while the remaining segments were labeled as quiet. Smoothing was required because the unsmoothed sporadic nature of the CI pulse train could result in the incorrect labeling of speech tokens as quiet segments. Figure 4.1 illustrates this sporadic nature. A zoomed-in section of the speech token "asa" is visible in the upper right corner of the figure. Within this focused time block, channels 13-17 are in a speech state. However, gaps, mostly evident in channel 13 (the top channel in the region of interest) and 16 (the second channel from the bottom of the zoomed-in image), exist during this segment. Without smoothing, these small sections of "off" pulses would incorrectly be labeled as quiet segments.

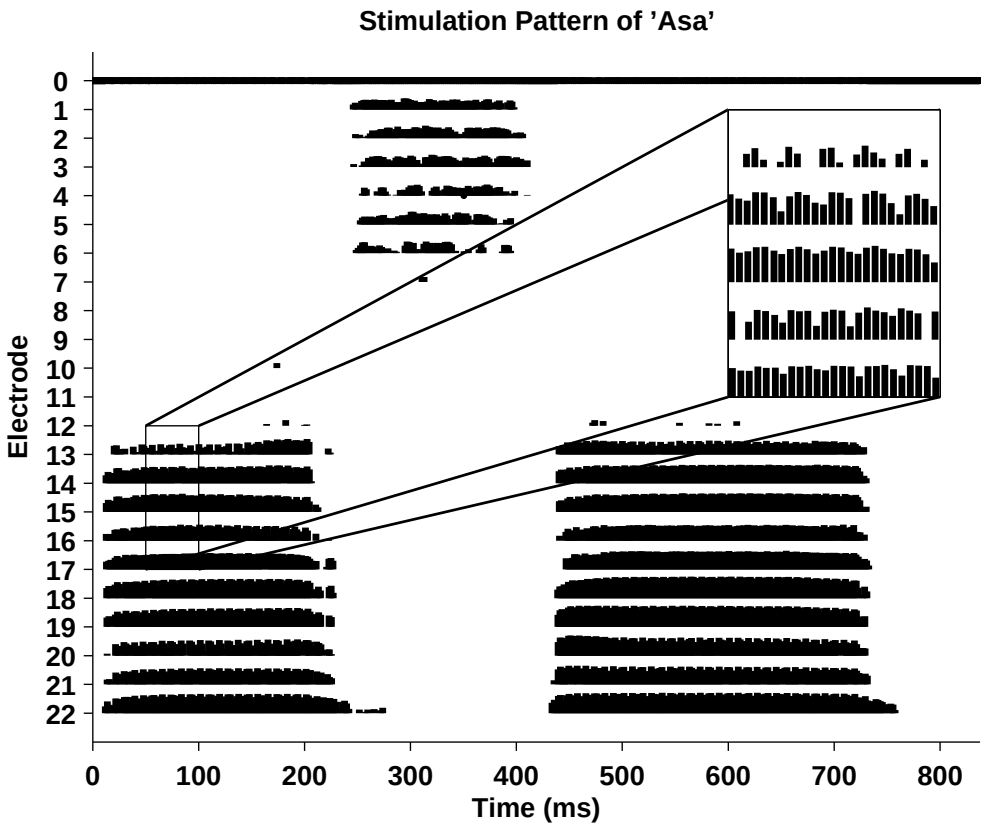

FIGURE 4.1: The speech token "asa" as processed by the ACE processing strategy. A section of the token has been magnified to visualize the sporadic nature of the CI pulse train. As can be seen in the zoomed-in portion of the figure, channels often alternate between an "on" and "off" state during speech segments. When labeling channel-specific stimuli as containing speech or quiet, smoothing is required to avoid mislabeling speech tokens as quiet segments during the sporadic "off" instances.

To isolate self-masking, effects of overlap-masking were removed from sentence stimuli by removing segments that were labeled as quiet in the reverberation-free speech but that contained stimuli in the reverberant speech. The resulting token contained solely self-masking effects. Conversely, to isolate overlap-masking, speech segments from the reverberation-free stimuli were inserted into the corresponding segments of the reverberant speech, resulting in stimuli containing only overlap-masking effects.

Figure 4.2 demonstrates an electrodogram containing labeled self- and overlap-

masking segments. Self-masking mitigation would be completed by inserting quiet speech tokens into the reverberant speech tokens shown in black, allowing the gray overlap-masking pulses to remain in the stimulus. Overlap-masking mitigation would require removing the gray pulses representing overlap-masking segments, while allowing the black reverberant speech segments to remain unaltered.

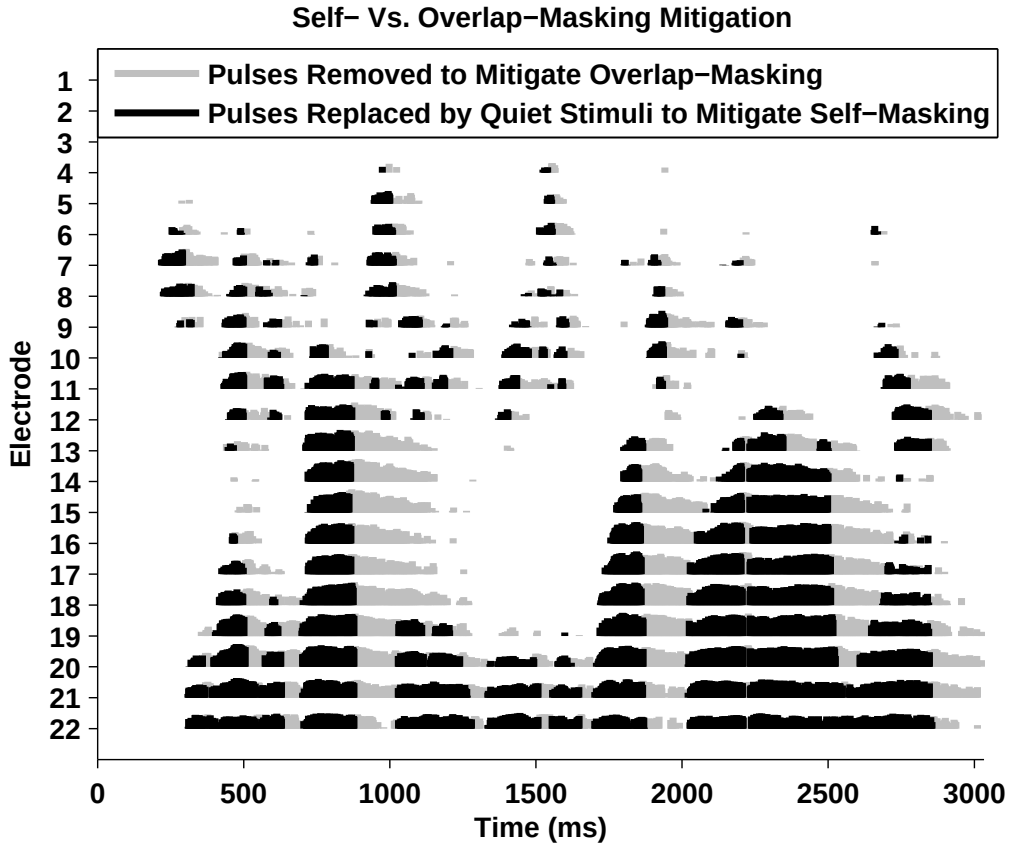

FIGURE 4.2: Electrodogram displaying both self-masking and overlap-masking effects. (The sentence shown here is "She had your dark suit in greasy wash water all year," from the TIMIT database.) Mitigating self-masking requires replacing the black pulses by quiet speech stimuli without altering the gray overlap-masking pulses. Conversely, to mitigate overlap-masking, the gray pulses would be removed from the token and the black pulses would remain unaltered [21].

Figure 4.3 displays a more detailed visual of ideal self- and overlap- masking mitigation. The top left section of the figure displays an anechoic segment extracted

from an example channel during a sentence token. The top right plot shows the given time segment in the presence of reverberation. Specifically, self-masking is visible in the amplitude corruption occurring in the black active speech segments, and overlap-masking is evident by the additional gray pulses that exist solely in the reverberant stimuli. Ideal overlap-masking mitigation (bottom left) causes the amplitude corruption during active speech to remain, while removing the gray overlap-masking pulses. Alternatively, ideal self-masking mitigation (bottom right) corrects the amplitude corruption in the black active speech segment, while allowing the overlap-masking pulses to remain.

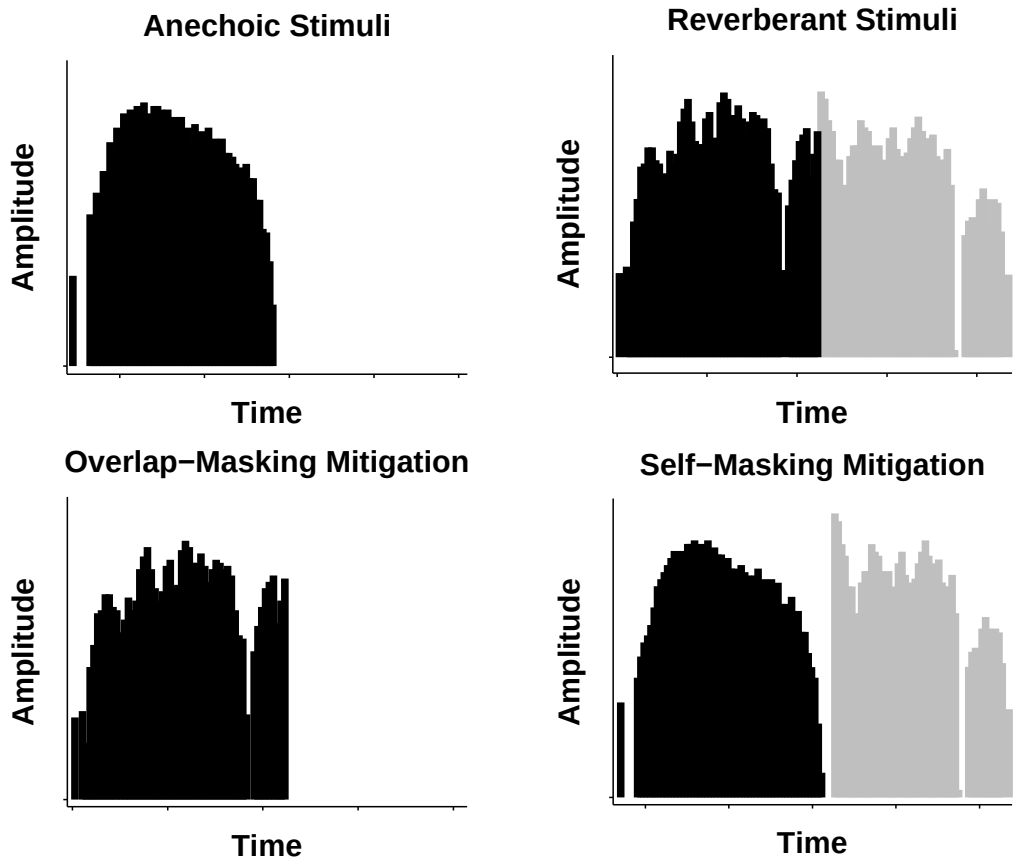

FIGURE 4.3: An example of the stimuli present on a given channel in quiet (top left), reverberation (top right), reverberation after ideal overlap-masking (bottom left), and reverberation after ideal self-masking mitigation (bottom right). After reverberation is added to the anechoic signal, self-masking effects are visible in the amplitude corruption occurring in the black active speech segment, and overlap-masking effects are visible in the additional gray pulses (top right). Ideal overlap-masking mitigation (bottom left) allows the amplitude corruption in the active speech segment to remain, but removes the overlap-masking pulses in gray. Conversely, ideal self-masking mitigation (bottom right) mitigates the amplitude corruption in the active speech segment, while having no effect on the overlap-masking pulses [21].

4.2.3 NH Methods

Ten normal hearing listeners were presented with speech using the acoustic model described in Section 4.1.1. The subjects were presented with three types of stimuli: reverberant stimuli, reverberant stimuli from which self-masking effects were

removed, and reverberant stimuli from which overlap-masking effects were removed. Reverberation times of 0.5s, 1.0s, and 1.5s were presented in increasing order of difficulty, and each subject was presented with one list containing ten sentences from the Hearing in Noise Test (HINT) database per condition [56]. Prior to beginning the experiment, subjects were presented with a list of 10 vocoded sentences in quiet, in order to become familiar with vocoded speech. The order of the tasks within each RT (original reverberant speech, reverberant speech after self-masking mitigation, and reverberant speech after overlap-masking mitigation) was randomized.

The percentage of words correctly identified per subject per condition was calculated. All words were considered in the final score, excluding articles. Only words that were completely accurate, with the exception of plurality, were scored as correct.

4.2.4 CI Methods

Four CI listeners were presented with the same three conditions as the NH subject group (reverberant stimuli, reverberant stimuli after self-masking mitigation, and reverberant speech after overlap-masking mitigation). Three lists (each containing ten sentences) were presented per condition, and the reverberation time was set to 1.5s. The Central Institute for the Deaf (CID) sentence database [61] was used, as the CI listeners had begun to learn the HINT sentences after performing other studies in our lab.

4.3 Results: Self- and Overlap- Masking Effects on Speech Intelligibility

Given the total number of correctly identified keywords and the number of keywords presented in all test sentences, a beta distribution was used to describe the probability of correctly identifying a keyword [62]. In the plots containing the results, displayed in Figures 4.4 and 4.5, the height of each bar represents the distribution's mean, and

95% confidence intervals are demonstrated by the error bars. Statistically significant differences in performance between conditions is illustrated by a line connecting the two corresponding bars. $(P_{condition_1} > P_{condition_2}) >= 0.95$ was used to determine statistical significance.

4.3.1 Normal Hearing Experiment

Figure 4.4 displays the results of the NH listening experiment pooled across subjects. The percentage of keywords correctly identified in the NH data (grouped for all subjects) is shown for unmitigated reverberation, reverberation after self-masking mitigation, and reverberation after overlap-masking mitigation. Statistically significant improvements in speech recognition resulted from mitigating either self- or overlap-masking at all three reverberation times. Further, mitigating overlap-masking resulted in statistically significant improvements in speech recognition compared to mitigating self-masking at reverberation times of 1.0s and 1.5s.

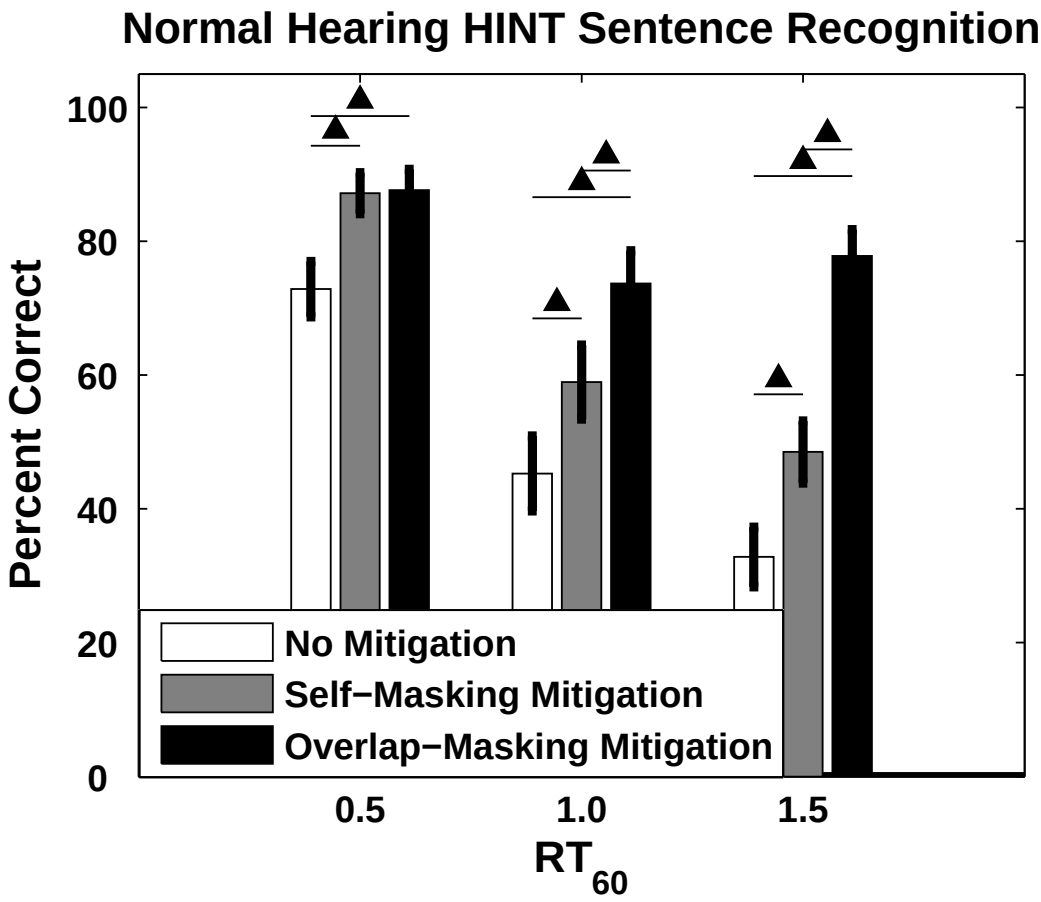

FIGURE 4.4: Speech recognition performance averaged across subjects for NH listeners presented with unmitigated reverberant speech and reverberant speech after ideal self- or ideal overlap- masking mitigation. The height of the each bar represents the mean of the performance distribution, and the error bars illustrate the 95% confidence intervals. Lines connecting two bars represent statistically significant differences. At all three reverberation times, speech recognition statistically significantly improved after mitigating either reverberation effect for the grouped data. Overlap-masking mitigated speech resulted in statistically significantly better speech recognition performance than self-masking mitigated speech at reverberation times of 1.0s and 1.5s [21].

4.3.2 Cochlear Implant Experiment

The speech recognition performance of four CI listeners presented with a reverberation time of 1.5s are shown in Figure 4.5. Each row provides the results for an individual subject. Reverberant speech after either self- or overlap- masking mitiga-

tion resulted in statistically significant improvements in speech recognition compared to unmitigated reverberant speech for all four subjects.

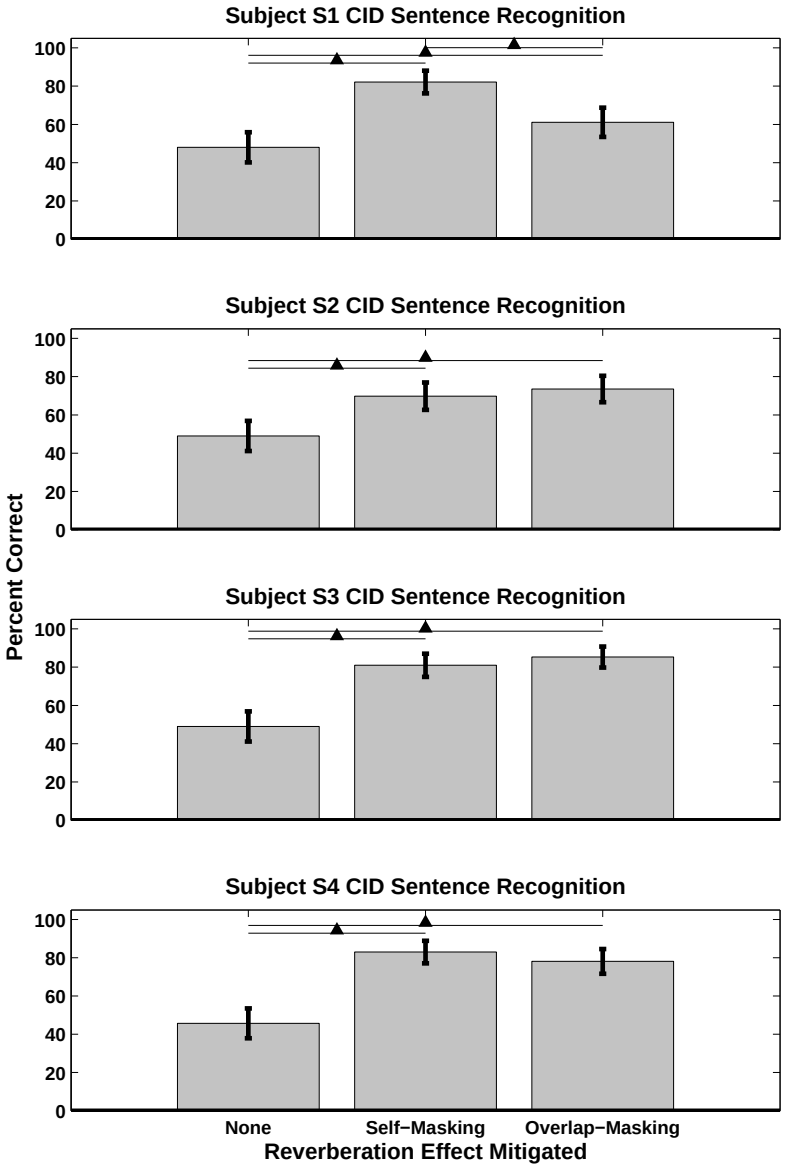

FIGURE 4.5: Speech recognition results shown for four CI subjects (down rows) in reverberant conditions with an RT of 1.5s after no reverberation mitigation, after ideal self-masking mitigation, or after ideal overlap-masking mitigation. Each bars' height represents the mean of the performance distribution, error bars signify 95% confidence intervals, and lines connecting two bars represent significant differences. Statistically significant improvements in speech recognition for all four subjects was seen after mitigating either self- or overlap- masking, when compared to unmitigated reverberant speech [21].

4.4 Discussion

This experiment suggests that mitigating either self-masking or overlap-masking has the potential to improve speech recognition for CI listeners in reverberant environments. Similarly, Kokkinakis and Loizou, 2011 found both self- and overlap- masking to be detrimental to speech intelligibility. However, their study found that self-masking may be the dominant cause of speech recognition degradation. Conversely, the current experiment found no statistically significant difference between the detrimental effects of the two forms of masking for three of four CI listeners. Further, the current experiment found that for two of the three reverberation times presented to the NH listeners, overlap-masking was statistically significantly more detrimental to speech recognition than self-masking. The discrepancy between the two studies' results may be explained by the different implementations of masking mitigation. While the current work referenced the reverberant and non-reverberant implant pulse trains to mitigate the reverberation effects ideally, Kokkinakis and Loizou, 2011 either replaced reverberant vowels with clean vowels (approximating self-masking mitigation) or replaced reverberant consonants with clean consonants (approximating overlap-masking mitigation) [1]. The assumption made by their study was that vowel stimuli fully describe self-masking effects while consonant stimuli fully contain overlap-masking effects. This study was therefore confounded by the influence of vowel and consonant information on speech intelligibility. Because a previous study found that vowels are more informative for speech intelligibility than consonants [63], the results of Kokkinakis and Loizou, 2011 may have been influenced by the inclusion of clean vowel information when considering self-masking effects and clean consonant information when considering overlap-masking effects. The current experiment, however, ideally mitigated reverberation without the confounding factors of the importance of vowel and consonant information for speech recognition.

Because the current experiment concluded that mitigating either self-masking or overlap-masking increased CI speech intelligibility in reverberant conditions, speech processing algorithms could benefit from a reverberation mitigation algorithm that mitigates either effect. Mitigating self-masking involves correcting amplitude corruption, which may require an estimation of the RIR. On the other hand, because overlap-masking exists after the speech signal has terminated, mitigating its effects can be accomplished by detecting and removing the associated pulses via machine learning techniques. Therefore, Chapter 5 describes the development of two overlap-masking mitigation strategies. One implementation focused on mitigating overlap-masking using an acoustic input signal, while another strategy utilized the CI pulse train with the goal of determining whether statistical models based on the limited information that is presented in the CI pulse train would be as efficacious as models based on higher resolution signals. The advantage of mitigation algorithms based on the CI pulse train is the possibility of more direct incorporation into the CI speech processing algorithm. To avoid interfering with a quiet speech processing strategy in non-reverberant conditions, it is envisioned that these overlap-masking mitigation strategies would be initiated by the reverberation detection algorithm outlined in Chapter 3.

5

Reverberation Mitigation

As demonstrated in Chapter 4, both self- and overlap- masking effects are significantly detrimental to CI speech intelligibility in reverberant environments. Mitigating self-masking effects requires that the amplitude corruption be corrected and may require knowledge of the RIR in order to inverse filter the reverberant signal. Such an algorithm may pose difficulties for real-time implementation. Alternatively, because overlap-masking occurs after active speech has terminated, mitigating its effects simply requires removing the pulses associated with overlap-masking. Therefore, machine learning techniques were used to detect and subsequently remove overlap-masking pulses from implant stimuli. Desmond et al., 2013 utilized the limited information present in CI stimuli to develop a reverberation detector that was relatively insensitive to room and CI stimulation parameters [20]. The results of that study suggested that using the simplified CI pulse train may result in performance advantages compared to using the more complex acoustic signal for overlap-masking detection as well. To investigate this hypothesis, two overlap-masking detection algorithms were investigated; one algorithm processes the CI pulse train, while another algorithm processes the acoustic signal.

Because overlap-masking may occur in one frequency bin while active speech is simultaneously presented in a separate frequency bin, overlap-masking detection and mitigation must occur on a channel-by-channel basis for both acoustic and CI pulse train stimuli. To successfully detect overlap-masking using either data type, features modeling the signal properties under quiet and reverberant conditions were first extracted, the true labels (speech stimuli vs. overlap-masking pulses) were determined for training, and a detection algorithm was trained and tested on the dataset.

5.1 CI Pulse Train Reverberation Mitigation

Initially, a reverberation mitigation strategy was developed that used features extracted from the frequency-time matrices generated by a CI speech processing algorithm. Because it was unknown whether the additional information available in an acoustic signal would be helpful for overlap-masking mitigation, the performance of the CI-based detector was subsequently compared to that of an acoustic-based strategy.

5.1.1 Feature Extraction

Features were selected to capture the differences between overlap-masking segments and active speech. Figure 5.1 shows an example of the pulsatile information within a given channel during a speech time window (black) followed by that during an overlap-masking time window (gray). During overlap-masking (gray), active speech energy sources are no longer present, resulting in an exponential decay with time [e.g. 2]. Conversely, as speech energy drives the signal during active speech tokens (black), the same decaying trend is not present in these segments. Therefore, in order to discriminate speech and overlap-masking, one feature described the energy present in windowed tokens, while another feature consisted of the differences in energies of

consecutive time windows. The frequency bins were determined by the ACE CI processing strategy, and the time windows were sliding windows containing a 30 ms duration. The sliding windows advanced by 2 ms, corresponding to the given within-channel CI stimulation rate, such that each pulse had its own set of features. The 30 ms duration was determined experimentally. These two energy-based features were developed to discriminate the exponentially decaying overlap-masking time segments from the speech segments that are driven by active speech energy.

Because the reverberation mitigation algorithm aimed to classify each pulse as "speech" or "overlap-masking," non-windowed, pulse-specific features were also considered. Specifically, because the amplitudes of stimuli within decaying overlap-masking segments may be smaller than those from active speech segments, pulse amplitudes were modeled statistically. To further capture the pulse-specific decaying property of overlap-masking stimuli, the differences in amplitudes of consecutive active pulses were also included as a feature.

Finally, the standard deviation was calculated for the previously described windowed data. Windows containing active speech, with constantly changing excitation energy, would most likely have higher standard deviations than slowly decaying overlap-masking windows. In an effort to include all of the available information for detection, features were extracted from CI stimuli prior to maxima selection in the ACE processing strategy. All features were scaled to have zero mean and unit variance prior to classification.

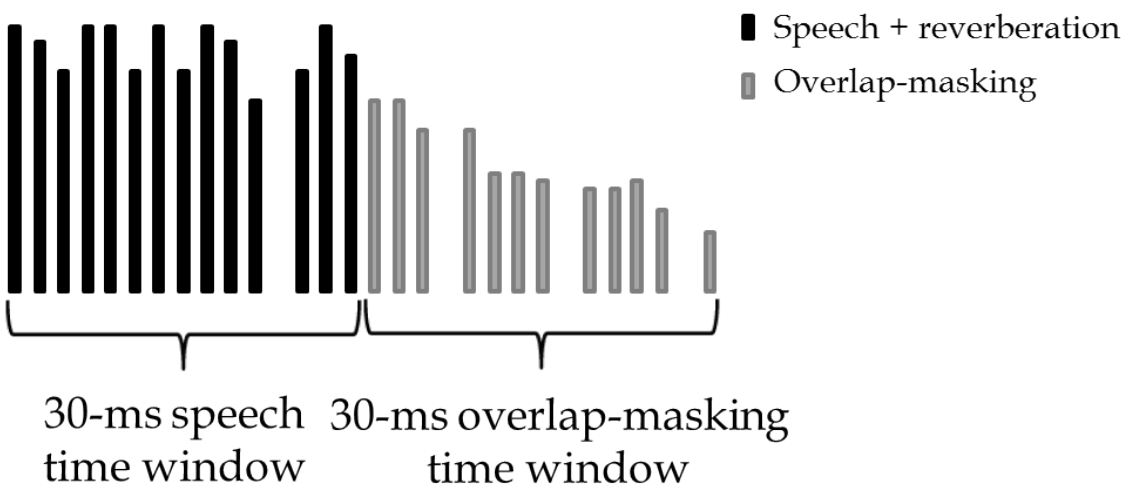

FIGURE 5.1: An example of a channel-specific 30 ms speech time window (black) followed by a 30 ms overlap-masking time window (gray). Because active speech energy sources are not present in overlap-masking segments, the corresponding pulse amplitudes decay exponentially with time [e.g. 2]. Conversely, the same decay is not present in active speech tokens, which result from active speech energy.

A plot of the probability density estimates of each classes' features is shown in Figure 5.2. (Recall that the features have been scaled to have zero mean and unit variance). The colors represent separate classes (blue and red represent speech and overlap-masking, respectively), and feature labels are shown along the x-axis. Data is displayed for a speech-related low frequency electrode (electrode 20). The windowed energy differences, energies, pulse amplitudes, and standard deviations are often larger for speech tokens compared to overlap-masking segments, which can be explained by the decaying nature of the latter's amplitudes. The differences between the amplitudes of the given pulse and the previous pulse, however, are more variable in speech tokens. Because the source is no longer active during overlap-masking, we expect a subtle decay between pulses, while the speech token amplitudes vary to a greater degree as the excitation energy changes. This variability is most likely not seen in the windowed-energy differences because, by considering data across a 30 ms time window, the overall trend is modeled, rather than the pulse-to-pulse variability.

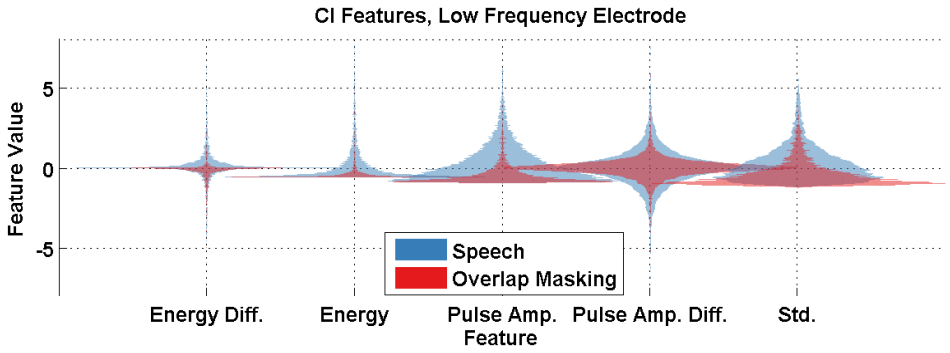

FIGURE 5.2: Probability density estimates of each feature within each class for CI overlap-masking detection. The features, shown here for data selected from one sentence, are labeled on the x-axis (from left to right: the difference in windowed-energies, the windowed-energy, the given pulse amplitude, the difference between the given pulse amplitude and the previous pulse amplitude, and the standard deviation). The two classes are shown in blue (speech) and red (overlap-masking). Although some separability is obvious here, the interaction between features, not illustrated here, often increases the between-class feature separability.

5.1.2 Labeling Truth

For training and testing purposes, the CI data must be labeled as "speech", "overlap-masking", or "quiet". The anechoic pulse train was referenced to create the truth labels, which was completed using methods similar to those described in Section 4.2.2. Specifically, the signal was smoothed using the same channel-specific summing filter spanning five time windows. Sums that were greater than each subjects' channel-specific thresholds were labeled as speech tokens, and the remaining segments were labeled as "quiet". Active pulses in the unsmoothed reverberant pulse train that occurred in the quiet segments of the anechoic signal were labeled as overlap-masking. Smoothing was not required for the reverberant pulse train because unactivated pulses were not of interest for classification since they would not affect a stimulation pattern.

5.1.3 Application of the RVM to CI Overlap-Masking Detection

A relevance vector machine (RVM) was implemented to detect overlap-masking within CI reverberant stimuli. As discussed in Section 3.2.2, RVMs are able to precisely describe data distributions by placing kernel functions at the feature locations. The less informative kernels are then pruned to create sparsity [58]. Because it was successful in detecting reverberation in general, as discussed in Chapter 3, this study hypothesized that the RVM would also be successful at detecting a specific effect of reverberation, overlap-masking. Other classifiers were considered, but when implemented resulted in poorer performance than the RVM.

5.1.4 CI Classifier Performance

In this example, the CI-based classifier was trained on ten sentences randomly selected from the TIMIT database, and testing was completed using a separate set of ten sentences. Reverberation with an RT of 1s was added to all sentences prior to processing. Because the overlap-masking mitigation algorithms will be implemented after reverberation has been detected in a pulse train, as discussed in Chapter 3, only speech in reverberation was considered for this study.

The results of the CI overlap-masking detection algorithm are shown for selected electrodes in Figure 5.3. The receiver operating characteristic (ROC) curves display the probabilities of detection as functions of the probabilities of false alarms. For application as an overlap-masking mitigation algorithm, a threshold must be selected that considers the trade-off between both probabilities.

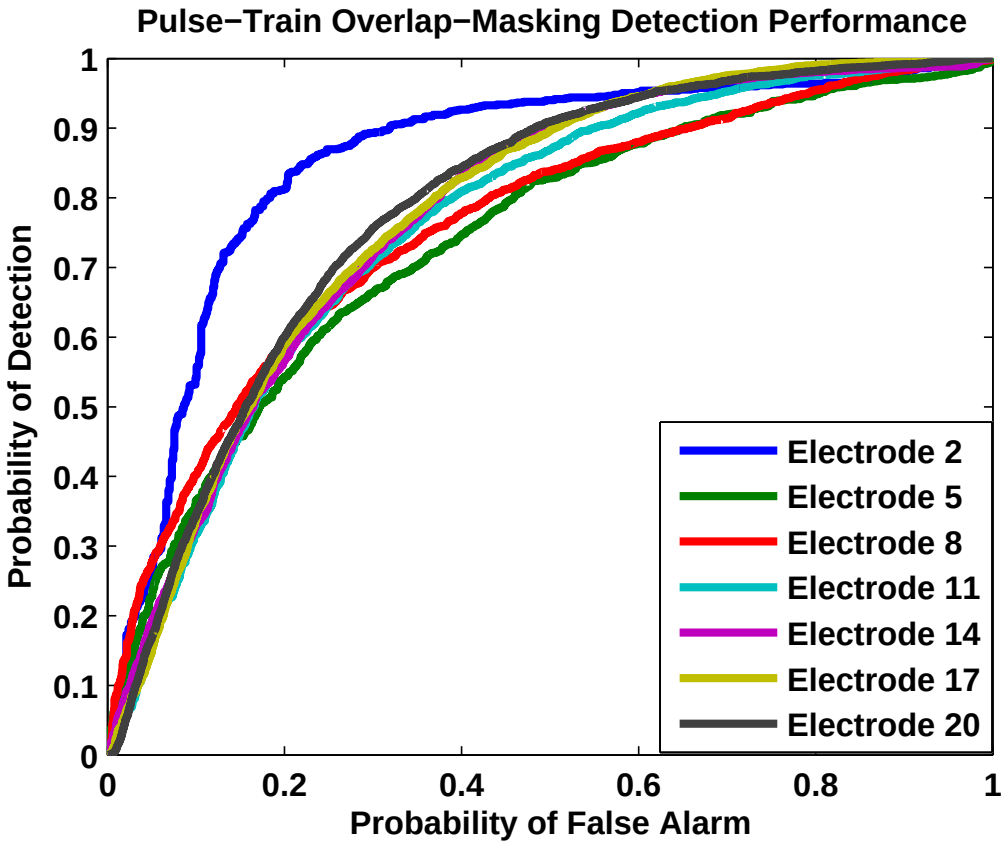

FIGURE 5.3: Performance of the CI overlap-masking detector, shown for electrodes 2, 5, 8, 11, 14, 17, and 20. The ROC curves display the probabilities of detection as functions of the probabilities of false alarm. A threshold that considers the trade-off between these probabilities will be selected for implementation.

Once implemented into a speech processing algorithm, correct detections will result in overlap-masking pulses being removed from the stimuli. False alarms, on the other hand, will result in the incorrect labeling of speech stimuli as overlap-masking, followed by their removal from the speech token. Although the results demonstrated in Figure 5.3 suggest that a high false alarm rate will be required to achieve an adequate probability of detection, this is not necessarily the case. It is possible that a low probability of detection, which would result in a lower false alarm rate, may be adequate for speech recognition improvements in reverberation, as some previously-masked information may become more audible. Alternatively, a

higher false alarm rate may be acceptable for speech perception. Because speech contains redundant information, many studies have found that CI listeners are able to compensate for dropped pulses [64; 65; 66; 67; 68; 69].

5.2 Acoustic Reverberation Mitigation

Similarly to CI overlap-masking detection, acoustic overlap-masking detection must be performed within time-frequency bins in order to align detection results with CI pulses. Because the acoustic overlap-masking mitigation algorithm will eventually be applied to mitigate reverberation in cochlear implants, the acoustic signal was initially bandpass filtered into logarithmically spaced frequency bands corresponding to those resulting from CI processing.

The general overlap-masking detection methods were also comparable for both the CI and acoustic implementations. Initially, features describing the different signal classes must be extracted from the speech stimuli, the true state of each signal at each moment in time (speech, overlap-masking, or quiet) must be labeled, and a detection algorithm must be trained and tested using the extracted features.

5.2.1 Feature Extraction

Initially, acoustic variations of the features used for CI overlap-masking detection were considered. The acoustic windowed features were calculated as described in Section 5.1.1, using 30 ms time windows. Because this detection algorithm will eventually be applied to CI data, sliding time windows advanced by 2 ms, corresponding to the pulse rate of the CI parameter set under consideration. This ensures that a unique set of features was calculated for each corresponding CI pulse.

However, the pulse-specific features (the pulse amplitude and the differences in amplitudes of consecutive active pulses) required calculations unique to the acoustic signal, which does not present pulsatile information. Using a pulse rate of 500pps,

one CI pulse corresponds to a 2ms acoustic time window. To ensure that each feature corresponded to one CI pulse, the acoustic versions of "pulse-specific features" were calculated using average amplitudes over 2ms time windows.

The probability density estimates of each acoustic classes' features (scaled to have zero mean and unit variance) are shown in Figure 5.4. Each color represents a different class (speech is shown in blue and overlap-masking is represented by red), and the feature labels are given on the x-axis. Features are shown for a low frequency bin (corresponding to electrode 20). Similarly to the CI-based features, the windowed energy differences, energies, pulse amplitudes, and standard deviations of the data are often larger for speech tokens compared to overlap-masking segments. Also comparable to the features extracted from CI stimuli, the differences between consecutive pulse amplitudes are more variable in speech tokens.

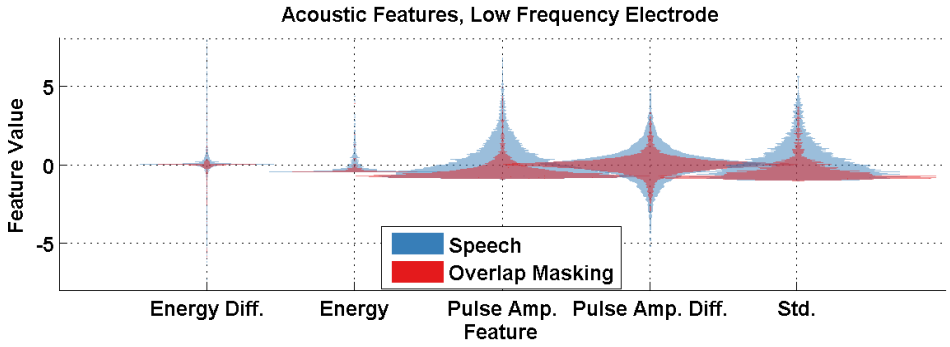

FIGURE 5.4: Probability density estimates of each feature within each class for acoustic overlap-masking detection. The features have been selected from an example sentence, and they are labeled on the x-axis (from left to right: difference between consecutive windowed-energies, windowed-energy, the average amplitude within a 2ms window, the difference in consecutive mean amplitudes calculated over 2ms windows, and the standard deviation). The classes are shown in blue (speech) and red (overlap-masking). The interaction between features, not shown here, may increase the between-class feature separability.

5.2.2 Labeling Truth

Because the acoustic overlap-masking detection algorithm will eventually be applied to mitigate reverberation in CI pulse trains, CI stimuli were used to label the acoustic truth matrices. Although methods were similar to those described in Section 5.1.2 for CI overlap-masking detection, slight modifications were required. Previously, the reverberant stimuli did not require smoothing, as off pulses were not of interest to the classification algorithm. However, acoustic stimuli is more continuous than CI processed stimuli, requiring both the reverberant stimuli and the anechoic stimuli to be smoothed prior to generating the truth matrices. For both stimuli, windows with sums less than the subject- and channel- specific thresholds, as specified in each CI listener's parameter set, were labeled as quiet, while those above the thresholds were labeled as containing stimuli. Segments that contained stimuli in the reverberant tokens but not the anechoic versions were labeled as overlap-masking, segments that contained stimuli in both tokens were labeled as speech, and the remaining segments were labeled as quiet.

5.2.3 Acoustic Classifier Performance

Because the RVM successfully detected overlap-masking in CI stimuli, it was hypothesized that the classifier may also effectively detect overlap-masking in acoustic stimuli. To explore this hypothesis, the RVM was applied to the acoustic features described in Section 5.2.1, and the results are shown in Figure 5.5. When using comparable features and the RVM, the acoustic classifier results in similar performance to the CI classifier, as expected. However, the potential advantage of using acoustic data is that more information may be available to improve classification. With that in mind, several additional features were investigated.

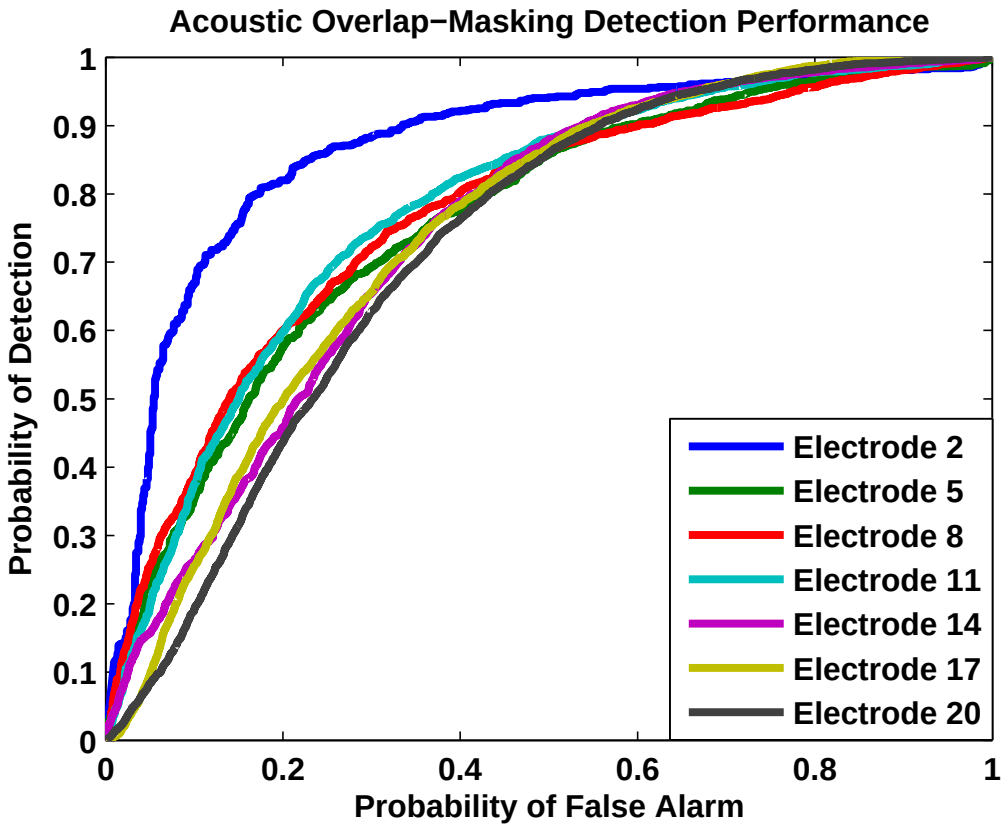

FIGURE 5.5: Overlap-masking detector performance resulting from the application of an RVM to the acoustic features. Performance is comparable to that obtained using CI features.

Effect of Additional Features on Performance

Although the application of an RVM to both the acoustic and CI-based features resulted in comparable overlap-masking detection accuracy, the addition of acoustic-specific features may further improve performance. Specifically, the average modulation depth within each time window was investigated as a feature because reverberation has been found to reduce this depth [70]. The correlation between two consecutive time windows was also considered because the reverberant signal present during overlap-masking was hypothesized to correlate with the previous active speech window. Additionally, features that utilize linear predictive coding (LPC) residuals were considered, as the LPC residuals of a reverberant signal approximate the con-

volution of an anechoic signal's LPC residuals with the RIR [51; 52; 32]. Specifically, the difference in consecutive time windows' LPC residual energies was investigated. Finally, to approximate the similarities between consecutive time windows, one window's LPC coefficients were applied to the following window, and the resulting residual energy was used as a feature.

Figure 5.6 displays the area under the curve (AUC) plots for the ROCs that resulted when each of these features was added to the base set of acoustic features described in Section 5.2.1. The AUCs are plotted as a function of electrode, and the CI overlap-masking detector AUCs are plotted in blue for reference. This figure shows that the addition of each acoustic feature independently did not improve acoustic performance above that of the CI overlap-masking detector. Additionally, including all acoustic features simultaneously (shown in black) did not improve detection accuracy.

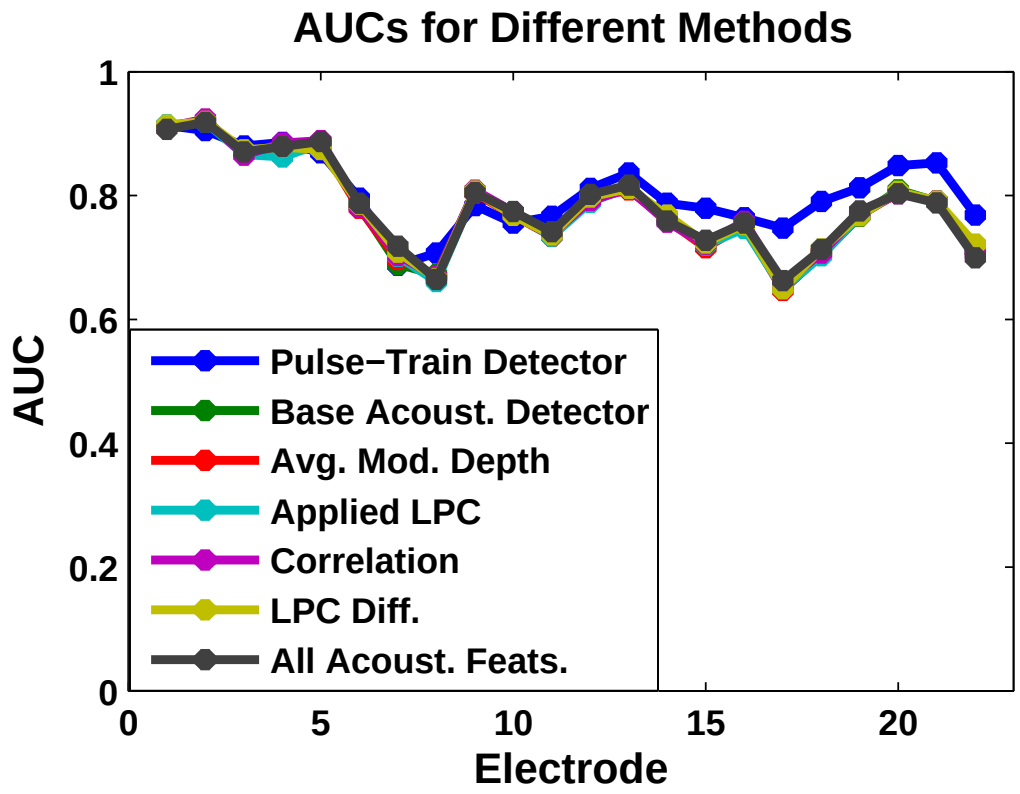

FIGURE 5.6: Area under the curve (AUC) plots for the ROCs resulting from the CI overlap-masking detector ("CI Detector"), the acoustic detector with similar features to the CI detector ("Base Acoust. Detector"), and the acoustic detector with the addition of features including the average modulation depth ("Avg. Mod. Depth"), the residual energy resulting from the application of one window's LPC coefficients on the following window ("Applied LPC"), the correlation between two time windows ("Correlation"), and the difference in LPC residual energies between consecutive time windows ("LPC Diff."). Classification performance using all acoustic features simultaneously is also shown ("All Acoust. Feats."). The addition of acoustic-specific features did not improve the acoustic detector performance.

While adding additional acoustic-based features did not result in improved detection, it was surprising that the acoustic-based features did not at least achieve the same performance as that provided by using pulse-based features. Similar performance was achieved between the pulse-based detectors and acoustic-based detectors for higher frequency channels, but not lower frequency channels. Because the band-pass filtered acoustic signals, unlike the CI channel-specific data, contain frequency

information within each band, it was hypothesized that lower frequencies may require longer time windows for better resolution. To test this theory, the AUCs resulting from the acoustic-based detector using two different feature windows is plotted in Figure 5.7. The pulse-based detector results are included for comparison. This plot suggests that a 30 ms time window benefits higher acoustic frequencies (lower numbered electrodes), while a 60 ms window performs better for lower acoustic frequencies. These results suggest that a frequency-dependent window may be best if acoustic-based features are used.

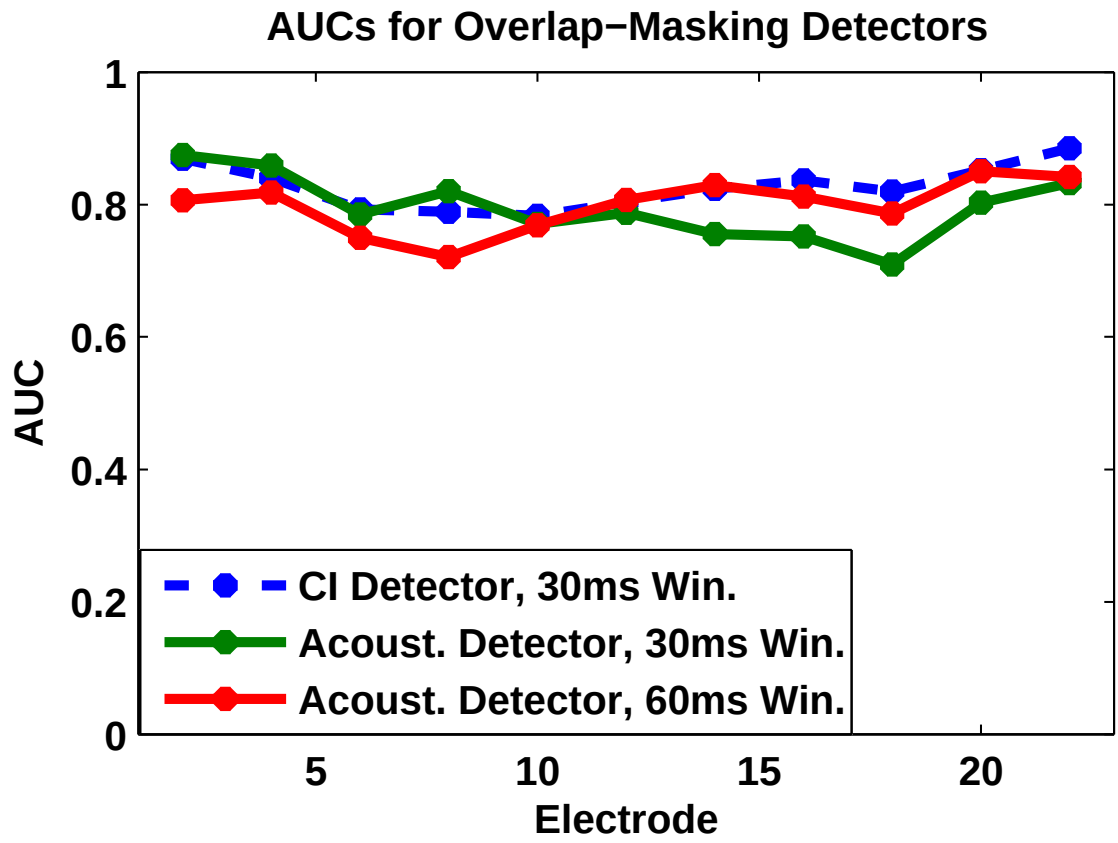

FIGURE 5.7: AUCs demonstrating overlap-masking detection performance for the CI detector ("CI Detector, 30ms Win.") compared to the acoustic detector using a 30ms time window for feature extraction ("Acoust. Detector, 30ms Window") and a 60ms time window ("Acoust. Detector, 60ms Window"). These results suggest that the acoustic detector performs better with a 30ms feature-extraction window at higher frequencies, while a 60ms window improves performance at lower frequencies (higher electrodes).

5.3 Discussion

Because overlap-masking occurs once the source signal is no longer active, removing its effects is as simple as accurately detecting its presence. Two methods, one utilizing the CI pulse train and another utilizing acoustic stimuli, were implemented for overlap-masking detection because it was unknown whether the limited information that is presented to a CI would be beneficial or detrimental for overlap-masking detection and mitigation. This study found that, with some probability of false

alarm, overlap-masking was detectable in various frequency channels using either CI or acoustic stimuli.

Although previous reverberation mitigation methods have been developed in the CI literature, no feasible real-time implementation exists. Hazrati and Loizou, 2013 implemented a reverberation mitigation strategy that thresholded on residual-to-reverberant ratios, but the algorithm required condition-specific parameter tuning [18]. Another study, completed by Hazrati et al., 2013, applied an adaptive threshold to a feature based on a ratio of variances, in an effort to mitigate overlap-masking effects. However, the algorithm assumes knowledge of the future signal, making real-time implementation difficult [19]. In a third example, Kokkinakis et al., 2011 implemented a channel-selection strategy based on the signal-to-reverberant ratio within each frequency channel for reverberation mitigation. Unfortunately, knowledge of the anechoic signal was required [7]. By utilizing only causal features, the current algorithm advances previous studies via the possibility of real-time implementation.

Overlap-masking was attempted acoustically using features similar to those used for CI detection. Using these descriptors and an RVM, detection performance was comparable to that resulting from the CI-based features. Because the addition of acoustic-specific features did not further improve performance, only the CI reverberation mitigation strategy was implemented into the CI speech processing algorithm. The efficacy of this algorithm at mitigating reverberation effects was evaluated in speech recognition experiments, as will be discussed in Chapter 6.

6

Reverberation Mitigation Algorithm Assessment

As presented in Chapter 5, overlap-masking detectors based both on CI and acoustic signals were developed. Mitigating overlap-masking, which exists after active speech has terminated, requires detecting and removing the corresponding pulses from the CI stimulation pattern.

Because performance was comparable between the CI- and acoustic-based overlap-masking detectors, only the CI-based detector was implemented into the speech processing strategy. The CI pulse-train-based detector not only requires less data than the acoustic detector for overlap-masking mitigation, but it is also easily integrated into the speech processing algorithm, as it utilizes a CI-processed signal.

Once the CI-based detector was implemented into the speech processing strategy, experiments were conducted to determine the algorithm's effect on speech intelligibility. Initial experiments were performed with normal hearing listeners and an acoustic model because NH subjects are more readily available than CI listeners and because their similarity in speech recognition performance allows for the pooling of their results for thorough statistical analysis. In addition to exploring the efficacy of the algorithm at mitigating overlap masking, the initial experiments were designed to

select an algorithm operating point (threshold), defining the trade-off between the probability of detection (P_D) and the probability of false alarm (P_{Fa}). In a final experiment, CI listeners were recruited to further investigate the effects of overlap-masking mitigation on speech recognition.

6.1 Normal Hearing Subject Sentence Recognition Performance in Simulated RIRs

The overarching goal of this study was to examine the false alarm errors that a CI listener can tolerate while benefiting from reverberation mitigation. An initial experiment presented NH listeners with an acoustic model representing CI pulse trains in unmitigated reverberation as well as in reverberation after overlap-masking mitigation. The results were utilized to determine the efficacy of the mitigation strategy and to determine the final algorithm operating points. These operating points define the balance between the amount of overlap-masking pulses correctly detected and the number of speech stimuli falsely labeled as containing overlap-masking.

6.1.1 Algorithm Performance

The overlap-masking mitigation strategy was trained on simulated rooms with reverberation times ranging from 0.5s to 1.5s, widths between 2m and 11m, lengths ranging from 5m to 20m, and heights varying between 2.2m and 3.2m. The initial experiment presented NH listeners with speech in reverberation in simulated RIRs with a room dimension of (10.0 x 6.6 x 3.0)m, a source location of (2.4835 x 2.0 x 1.8)m, and a microphone located at (6.5 x 3.8 x 1.8)m [59]. Three reverberation times of 0.5s, 1.0s, and 1.5s were presented. Performance of the overlap-masking detection algorithm when applied to these simulated rooms is shown in Figure 6.1. From left to right, the figure displays performance ROCs for simulated RTs of 0.5s,

1.0s, and 1.5s. Within each plot, performance for electrodes 2, 5, 8, 11, 14, 17, and 20 is shown.

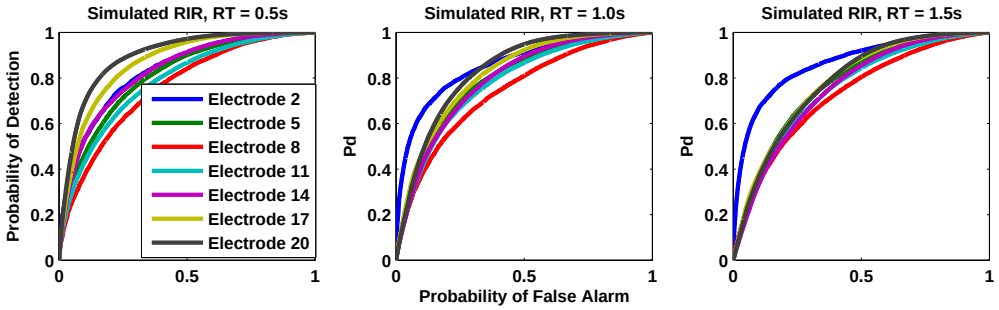

FIGURE 6.1: ROCs demonstrating overlap-masking detection performance in simulated RIRs with a room dimension of (10.0 x 6.6 x 3.0)m, a source location of (2.4835 x 2.0 x 1.8)m, and a microphone located at (6.5 x 3.8 x 1.8)m [59]. RT values were set to 0.5s (left), 1.0s (middle), and 1.5s (right). This figure demonstrates performance for the RIRs that were presented to NH listeners in an initial experiment.

6.1.2 Methods

Fifteen NH subjects were recruited to study the effects of reverberation mitigation on speech recognition. In order to become familiar with the CI acoustic model, subjects completed a training task prior to testing. During training, listeners were presented with twenty sentences, post-CI processing, from the TIMIT database [30], and they were instructed to type what they heard. Feedback was provided. Once the initial twenty sentences were presented, sentences continued to be presented until the subject's speech recognition plateaued, improving by less than 10% between consecutive five-sentence groups. The responses were graded automatically.

After the training session was complete, one list containing ten sentences from the HINT sentence database [56] was presented per subject per condition. The three reverberation conditions that were added to the speech signals prior to mitigation were those outlined in Section 6.1.1. The reverberation mitigation strategy was applied to the stimuli using channel-specific P_Ds of 90%, 75%, 60%, 45%, 30% and 15%, presented in random order. Speech intelligibility resulting from the applica-

tion of the mitigation algorithm was compared to that resulting from unmitigated reverberation.

6.1.3 Results

The NH speech recognition performance is shown in Figure 6.2. As discussed in Section 4.3, a beta distribution describes the probability of correctly identifying a keyword, given the total number of correctly labeled keywords and the number of keywords contained in all sentences. The distribution's mean is represented by the height of each bar, and error bars demonstrate the 95% confidence intervals. A line connecting two error bars illustrates a statistically significant difference in performance, and statistical significance is determined by $(P_{condition_1} > P_{condition_2}) >= 0.95$.

The three separate groups in Figure 6.2 represent performance in reverberation with RTs of 0.5s, 1.0s, and 1.5s. Within each group, performance in unmitigated reverberation is displayed in the left-most blue bar. Viewing results from left to right within each group, performance in mitigated reverberation is demonstrated for P_Ds of 15%, 30%, 45%, 60%, 75%, and 90%.

Speech recognition improved in mitigated reverberation using assorted thresholds. As the P_D increased from a value of 0 (unmitigated reverberation), speech recognition performance increased as more overlap-masking pulses were successfully removed. However, as P_D was further increased, speech intelligibility eventually began to decrease, as too many speech pulses were misclassified as overlap-masking and were therefore removed from the stimuli (i.e. as P_{Fa} increased).

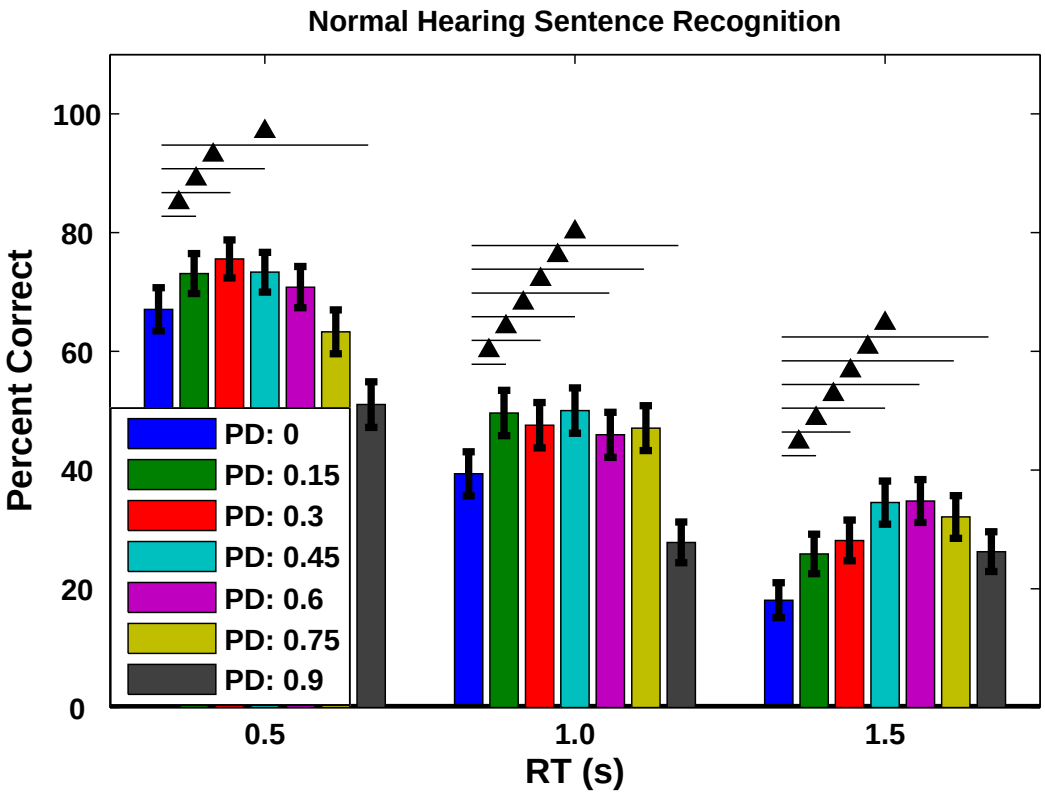

FIGURE 6.2: Speech recognition results for NH listeners using an acoustic model in simulated reverberation with RTs of 0.5s (left), 1.0s (middle), and 1.5s (right) in both unmitigated and mitigated reverberation. Each cluster represents performance for a different RT. Within each group, listeners were presented with unmitigated reverberation (blue: far left), and mitigated reverberation using various P_Ds (increasing in value from left to right).

One goal of this experiment was to determine the algorithm operating point(s) defining the trade-off between P_D and P_{Fa}. As previously mentioned, speech intelligibility initially improved with increasing P_D, followed by a decrease in performance as a greater P_D also resulted in a larger P_{Fa}. Based on the speech recognition results shown in Figure 6.2, it was determined that $P_D s$ of 30%, 45%, and 60% removed adequate overlap-masking stimuli, while simultaneously retaining enough speech pulses for improved speech recognition after reverberation mitigation. These thresholds, therefore, were selected for future experiments.

Although Figure 6.2 suggests that overlap-masking mitigation successfully improves speech recognition at varying RTs and varying P_Ds, this experiment was conducted using simulated RIRs and no added noise. A more real-world environment, however, might consist of both added noise and reverberation resulting from recorded RIRs. Therefore, Section 6.2.1 explores the effects of both real-world RIRs as well as added noise on speech recognition, both with and without reverberation mitigation.

6.2 Normal Hearing Subject Sentence Recognition in the Presence of Noise and Recorded RIRs

To determine the algorithm's robustness to noise, speech in the presence of speech-shaped noise (SSN) and reverberation as well as multi-talker babble and reverberation were presented to the subjects. Both SSN and multi-talker babble have been used in the literature to test CI speech recognition in noise [e.g. 71]. Reverberation was added to the signals using recorded RIRs, with the goal of creating a more realistic listening environment.

6.2.1 Algorithm Performance in Noise

As described in Section 6.1.1, training was completed using simulated reverberant rooms with RTs varying from 0.5s to 1.5s, widths between 2m and 11m, lengths ranging from 5m to 20m, and heights existing in the range of 2.2m to 3.2m. In this experiment, NH listeners were presented with RIRs recorded in a lecture hall (10.8m by 10.9m), an office (5.0m by 6.4m), and a corridor (18.25m by 2.0m), available from [72]. Before reverberation was introduced to the signal, sentences were either presented in quiet, corrupted with SSN, or corrupted with multi-talker babble.

Overlap-masking detection performance in the three recorded RIRs used in this experiment is shown across the columns of Figure 6.3, and performance in the three

types of added noise (none, SSN, multi-talker babble) is shown across the rows. Performance is displayed for electrodes 2, 5, 8, 11, 14, 17, and 20. This figure demonstrates that the mitigation algorithm trained in simulated RIRs with no added noise is robust to both recorded RIRs as well as to outside noise sources.

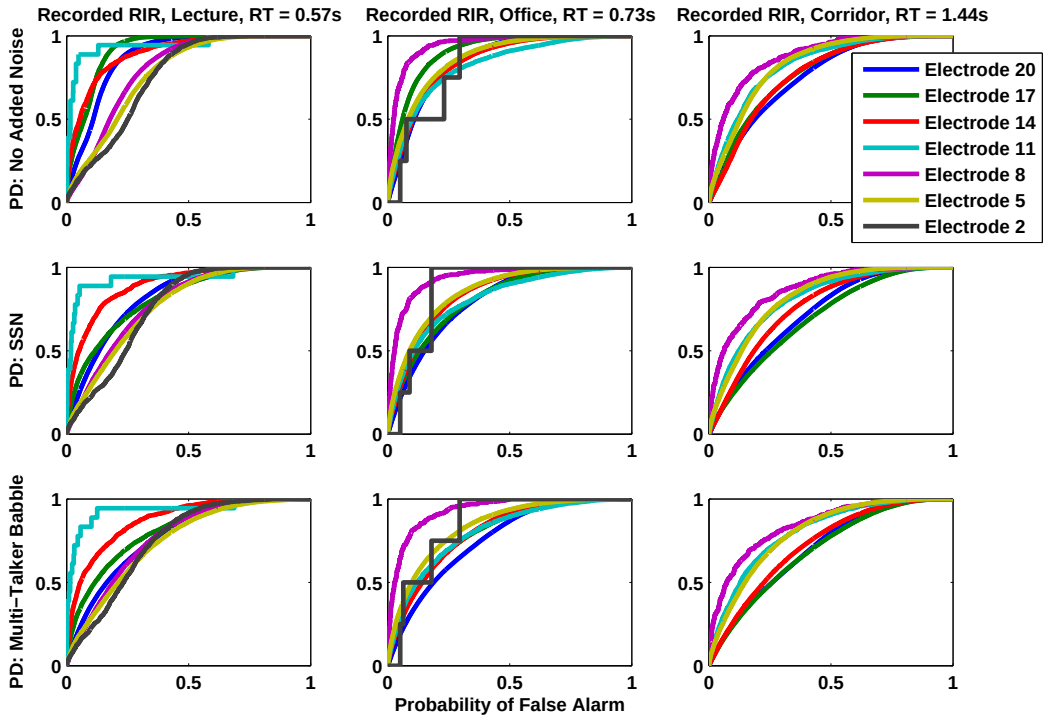

FIGURE 6.3: Overlap-masking detection performance in a lecture hall (left column), an office (middle column), and a corridor (right column), with either no added noise (top row), the addition of SSN (middle row), or the addition of multi-talker babble (bottom row). Performance is displayed in the form of ROCs for electrodes 2, 5, 8, 11, 14, 17, and 20. These results suggest that the reverberation mitigation strategy is robust to both recorded RIRs and to added noise sources.

6.2.2 Methods

NH subjects were recruited to investigate the effects of both recorded RIRs and added noise on speech recognition in unmitigated and mitigated reverberation. Similarly to the experiment described in Section 6.1.2, listeners initially completed a training task. This task was considered complete when speech recognition of sentences from

the TIMIT database [30] improved by less than 10% between consecutive groups containing five-sentences.

Subjects were then presented with speech in reverberation, speech in SSN and reverberation, or speech in multi-talker babble and reverberation. RIRs recorded in a lecture hall (RT = 0.57s), an office (RT = 0.73s), and a corridor (RT = 1.44s) were used [72], and unmitigated reverberant speech as well as speech in mitigated reverberation with P_Ds set to 0.3, 0.45, or 0.6 were presented.

The City University of New York (CUNY) sentence database was used for this experiment. Because additional experimental conditions were included to explore the effects of added noise and reverberation on speech recognition, a more extensive database than that provided by HINT was required. The CUNY database was selected as it not only contains enough sentences for the current experiment, but it also consists of "everyday sentences," similar to the HINT collection. One sentence list from the CUNY database was presented per subject per condition, and results were pooled across listeners.

In total, four NH studies were conducted, and the differences between the studies are highlighted in Table 6.1. The first difference was the type of RIR that was convolved with the stimuli, either simulated or recorded. The experiment introduced in Section 6.1 utilized simulated RIRs, while the experiments discussed in Sections 6.2.3, 6.2.4, and 6.2.5 presented stimuli in recorded RIRs. Next, the sentence databases varied. The initial experiment presented in Section 6.1 utilized professional recordings from the HINT database. The subsequent experiment, which will be introduced in Section 6.2.3, presented listeners with sentences that were recorded by seven male native English speakers at Duke University. Because the lists consisted of twelve sentences, each speaker recorded one or two sentences for each list. Finally, to ensure that non-professional recordings were not interfering with performance, professional recordings of the CUNY sentences were used in the experiments presented in Sections

6.2.4 and 6.2.5.

The final parameter that was adjusted between studies altered the method of algorithm threshold selection. In all experiments, electrode-specific thresholds were selected to achieve a given P_D. However, as shown in the performance ROCs plotted in Figure 6.3, algorithm accuracy varies in different noise and reverberation conditions. As a result of this variation, a threshold that achieves one P_D in a certain RIR may produce a different P_D in a separate RIR. Therefore, the accuracy of the target P_D is affected by whether thresholds are selected with or without knowledge of the reverberation parameters. Specifically, thresholds that are selected with knowledge of the RIR should result in more accurate experiential P_Ds than those selected without knowledge of the RIR.

Table 6.1: Parameters Affecting the NH Studies

Study Number	RIR Type	Sentence Database	Threshold Selection Method
1	Simulated	Professionally Recorded HINT	Known RIR
2	Recorded	Duke University CUNY	Unknown RIR
3	Recorded	Professionally Recorded CUNY	Unknown RIR
4	Recorded	Professionally Recorded CUNY	Known RIR

6.2.3 Results using Duke University CUNY Recordings and Unknown RIR Parameters for Threshold Selection

The study outlined in this section corresponds to Study 2 in Table 6.1. Sentences from the CUNY database, as recorded by Duke University, were exposed to recorded reverberant conditions. Thresholds were selected assuming no knowledge of the experimental RIR, and SSN and multi-talker babble were added prior to reverberation, with an SNR of 5dB.

Ten normal hearing listeners were recruited for this experiment, and the sen-

tence recognition results are presented in Figure 6.4. The top plot displays results in reverberation alone, the middle plot presents speech recognition in SSN and reverberation and the bottom plot shows performance in multi-talker babble and reverberation. Reverberation mitigation did not consistently improve speech recognition for all subjects across conditions in this experiment.

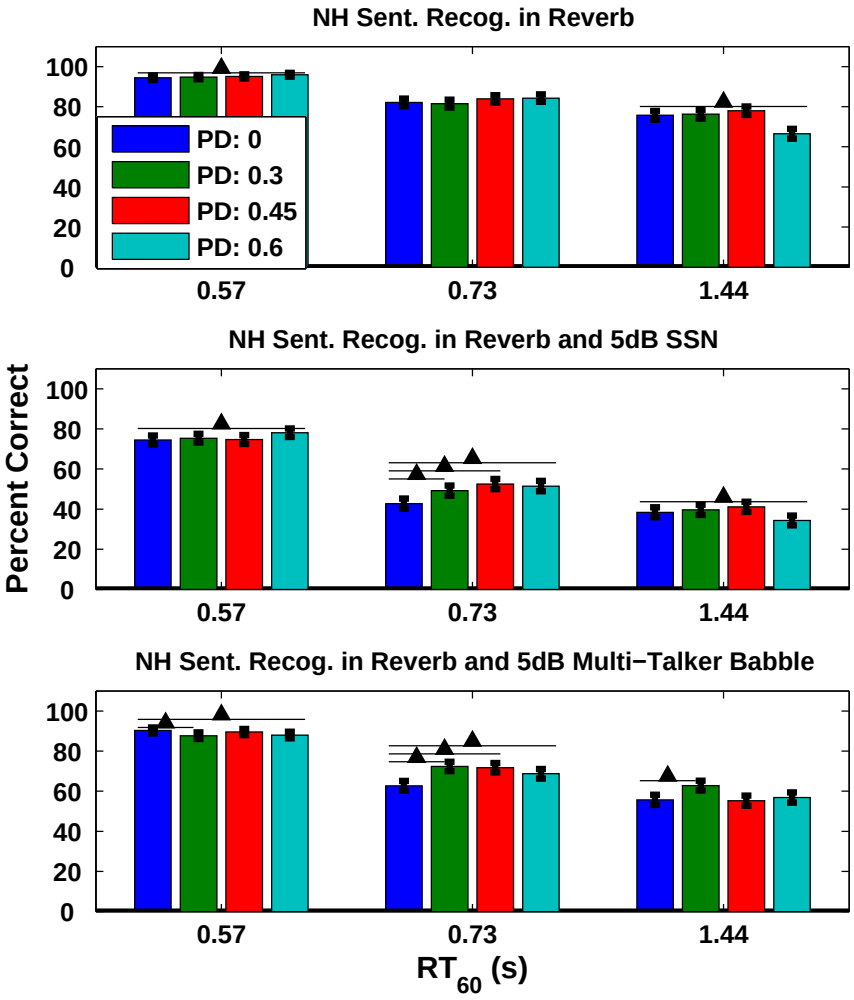

FIGURE 6.4: NH speech recognition performance in reverberation (top row), SSN with an SNR of 5dB and reverberation (middle row), and multi-talker babble with an SNR of 5dB and reverberation (bottom row). Within each plot, the first group of results was collected using the RIR recorded in a lecture hall, the middle group utilized an RIR recorded in an office, and the rightmost group used an RIR recorded in a corridor [72]. Speech recognition was studied in unmitigated reverberation (blue), as well as reverberation after overlap-masking mitigation at varying P_Ds. Reverberation mitigation did not consistently improve speech recognition for all subjects across conditions in this experiment.

As presented in Section 6.1.3, reverberation mitigation improved speech recognition performance across listeners in simulated RIRs. However, as shown in Figure

6.4, the same improvement was not observed in recorded rooms. In addition to differences between the RIRs used in each experiment, the sentence material differed between the two trials. The initial task utilized professionally recorded HINT sentences [56], while the experiment presented in this section used amateur recordings of CUNY sentences [73]. Because the HINT sentence database does not contain enough lists to run the current task, professional recordings of the CUNY sentences were obtained for a follow-up study. A preliminary experiment concluded that 8dB SNR should be added to the signals when presenting noise and reverberation simultaneously. Aside from the new sentence material and noise levels, the methods for this task remain identical to those outlined in Section 6.2.2.

6.2.4 Results using Professional CUNY Recordings and Unknown RIR Parameters for Threshold Selection

The results outlined in this section utilized the parameters given in Study 3 in Table 6.1. Fifteen NH listeners were presented with professionally recorded CUNY sentences, and their performance is displayed in Figure 6.5. Data is presented in a similar format to that shown in Figure 6.4, with sentence recognition in reverberation represented by the top plot, in SSN and reverberation displayed in the middle plot, and in multi-talker babble and reverberation shown in the bottom plot. Each subgroup represents data collected in different reverberation conditions. Results for unmitigated reverberation are displayed in blue, while the remaining colors represent performance after reverberation mitigation at varying P_Ds. Similarly to the results presented in Figure 6.4, speech recognition did not consistently improve after the application of the reverberation mitigation strategy.

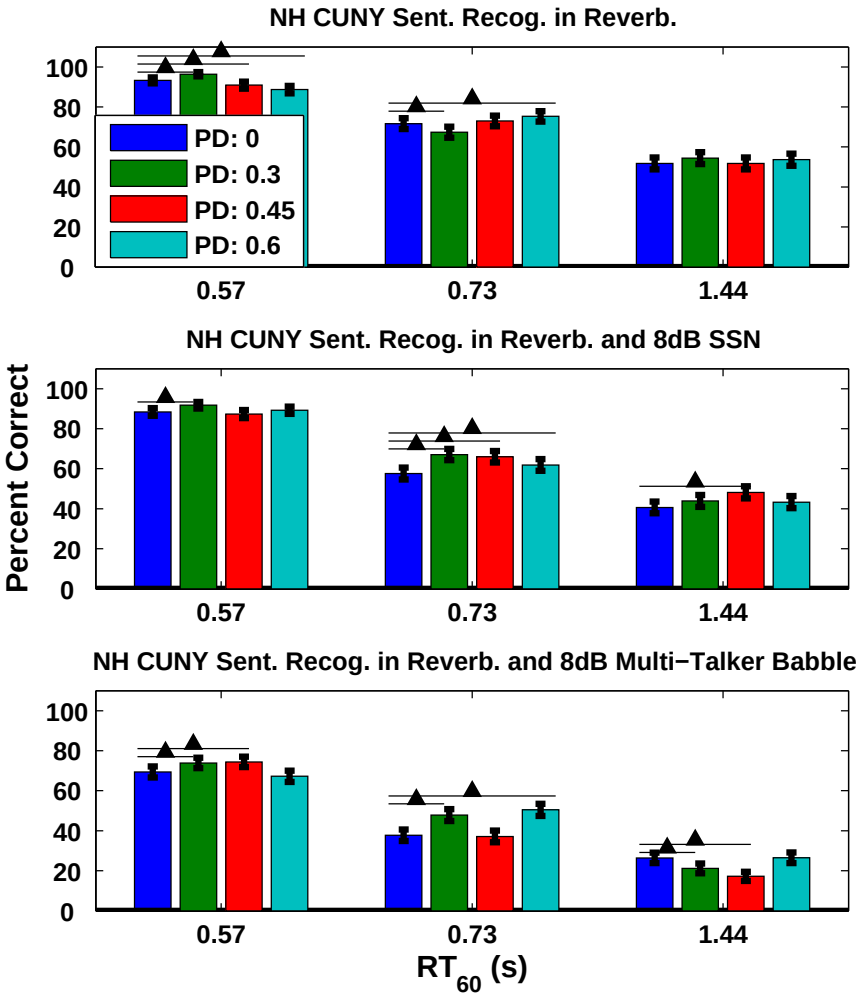

FIGURE 6.5: NH speech recognition performance using professionally recorded CUNY sentences in reverberation (top), SSN and reverberation (middle), and multi-talker babble and reverberation (bottom). RIRs were recorded in a lecture hall (left), an office (middle), and a corridor (right). Similarly to the results shown in Figure 6.4, performance did not consistently improve after the application of reverberation mitigation for all subjects and all conditions.

One final difference between the experiments conducted with recorded RIRs and that in Section 6.1.3 was the method of determining electrode-specific algorithm thresholds. The detection performance ROCs, demonstrated in Figure 6.3, vary in different noise and reverberation conditions. Therefore, a threshold that achieves one

P_D in a given RIR may result in a completely different P_D in a separate reverberant condition. When conducting the experiment in Section 6.1.3, thresholds were selected assuming knowledge of the reverberant room parameters, resulting in highly accurate P_Ds. Conversely, the experiments in this section did not assume such knowledge, resulting in P_Ds that may vary from the target values.

Figures 6.6 and 6.7 demonstrate this phenomenon. Both figures plot different recorded reverberant rooms along the columns (from left to right: a corridor, lecture hall, and office), while the different noise conditions are plotted across the rows (from top to bottom: reverberation alone, reverberation and SSN, reverberation and multi-talker babble). Each line within a given plot represents the data from one electrode. Figure 6.6 plots the electrode-specific kernel density estimates (KDEs) that would result if thresholds aiming to achieve a P_D of 45% were selected assuming knowledge of the reverberant conditions. As expected, most KDEs are centered around a P_D of 45%. Alternatively, Figure 6.7 displays the electrode-specific P_D kernel density estimates that result from thresholds selected without any knowledge of reverberation parameters. Compared to the KDEs in Figure 6.6, the distributions in Figure 6.7 vary much more, suggesting that the target P_Ds may differ significantly from the actual P_Ds that are presented to the listeners. As a result, one hypothesis was that the speech recognition performance presented in Figures 6.4 and 6.5 suffered as a result of inaccurate threshold selection.

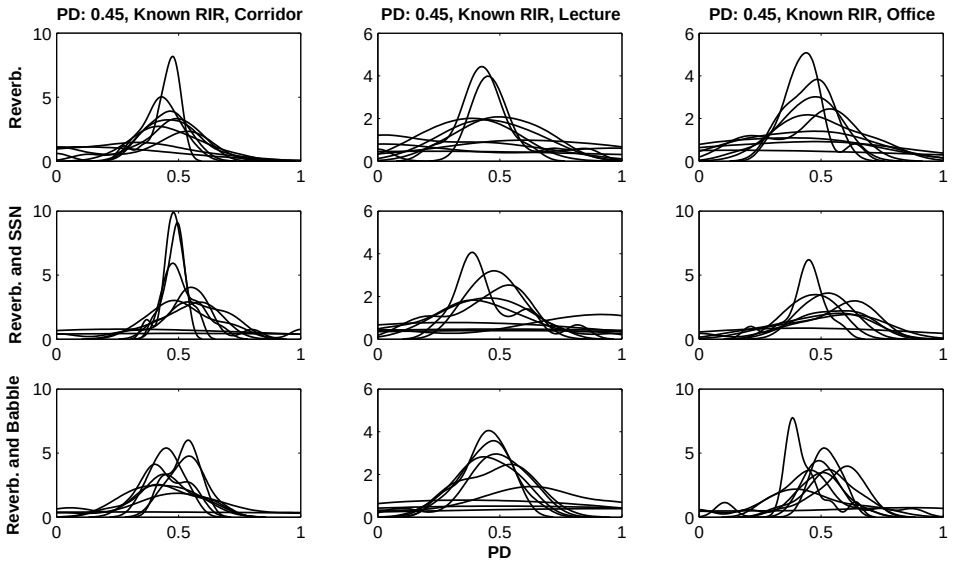

FIGURE 6.6: Electrode-specific kernel density estimates of experimental P_Ds. The thresholds used to achieve these P_Ds were determined with knowledge of the RIRs, and the target P_D was 45%. From left to right, the plots down the columns represent results in a corridor, a lecture, and an office. From top to bottom, the plots across the rows represent data collected in reverberation, reverberation and SSN, and reverberation and multi-talker babble. Because knowledge of the RIR was assumed, the KDEs are mostly centered around a P_D of 45%.

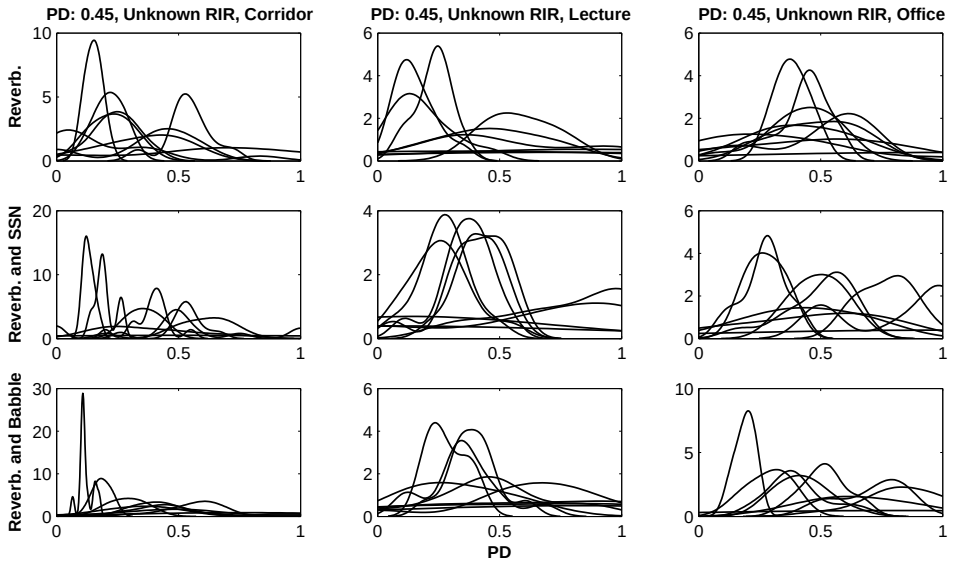

FIGURE 6.7: Electrode-specific kernel density estimates of experimental P_Ds. The thresholds used to achieve these P_Ds were determined without knowledge of the RIRs, and the target P_D was 45%. The plots down the columns represent tokens in a corridor, a lecture, and an office. The plots across the rows correspond to stimuli in reverberation, reverberation and SSN, and reverberation and multi-talker babble. The KDEs, which demonstrate P_Ds that result from thresholds selected without knowledge of the RIR, contain greater variation than those plotted in Figure 6.6, which result from threshold selection in known RIRs.

6.2.5 Results using Professional CUNY Recordings and Known RIR Parameters for Threshold Selection

To determine the effect of threshold selection on speech recognition in mitigated reverberation, a subsequent study was conducted. The study presented in this section corresponds to Study 4 in Table 6.1. Eleven NH listeners were presented with speech corrupted by recorded RIRs, both with- and without- reverberation mitigation. Thresholds for mitigation were calculated assuming knowledge of the reverberant conditions. Experimental methods are as outlined in Section 6.2.2, and the resulting speech recognition performance is demonstrated in Figure 6.8. As demonstrated in this figure, reverberation mitigation did not consistently improve speech recognition when assuming knowledge of the listening conditions for algorithm

threshold selection.

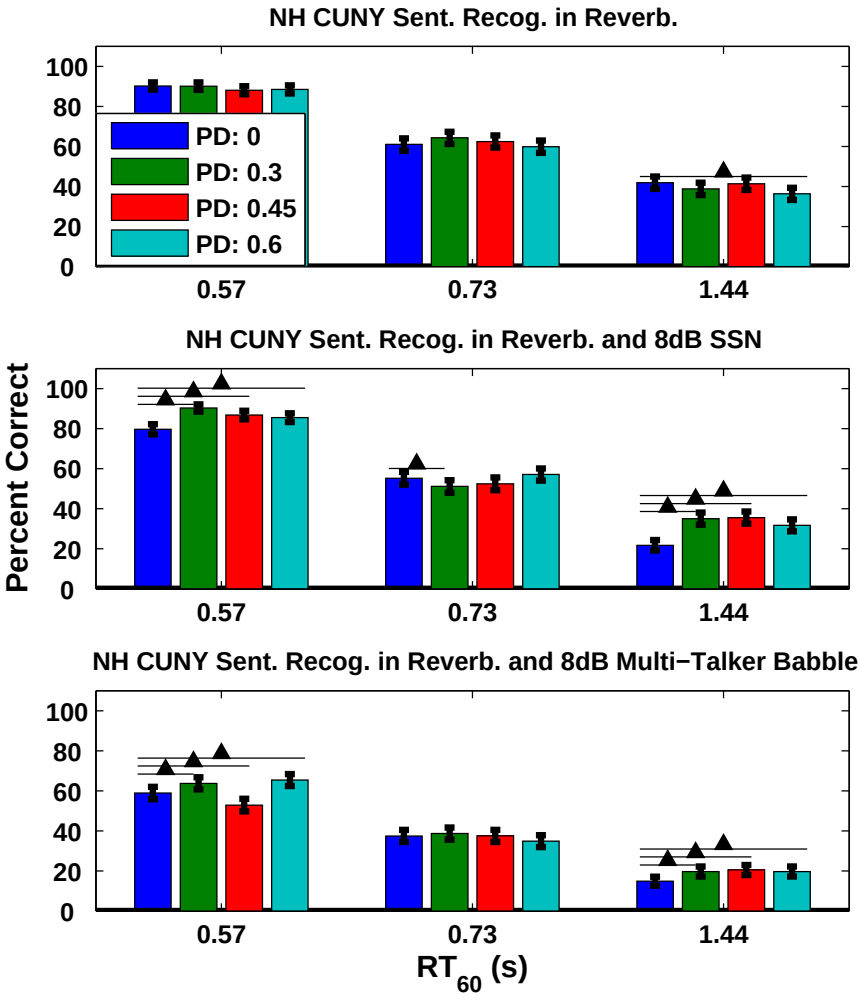

FIGURE 6.8: NH speech recognition performance in reverberant and noisy-reverberant listening conditions, with and without reverberation mitigation. The mitigation strategy applied in this experiment assumed knowledge of the RIR when selecting an operating point. Similarly to the results presented in Figures 6.4 and 6.5, speech recognition performance did not increase as a result of reverberation mitigation in recorded RIRs.

The reverberation mitigation strategy increased speech recognition performance in simulated reverberant environments, but was unable to improve speech intelligibility consistently for all subjects in recorded reverberant conditions. It was there-

fore hypothesized that the discrepancies between the performances in simulated and recorded reverberant rooms may be due to some fundamental difference between the two listening conditions.

6.3 Comparison of Simulated and Recorded RIRs

A visualization of a sentence in both a simulated and recorded environment is shown in Figure 6.9. These plots demonstrate sections of active speech in black and sections of overlap-masking in gray. For simplification, the stimuli in Figure 6.9 simply display binary pulses representing activity, and no amplitude information is conveyed.

The leftmost plot demonstrates the speech token in a simulated reverberant condition while the rightmost stimulus was exposed to a recorded RIR. Although there are comparable amounts of overlap-masking in the high electrode channels (low frequencies) between the two stimuli, the amount of overlap masking greatly reduces in the low electrode numbers (high frequency channels) in the recorded RIR condition. Conversely, the amount of overlap-masking activity remains relatively consistent across channels in the simulated environment.

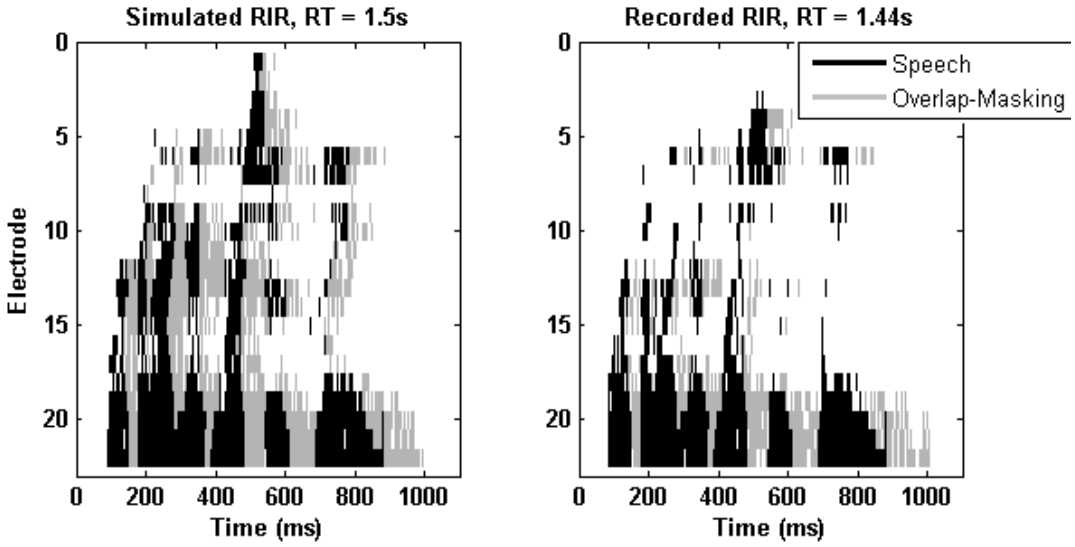

FIGURE 6.9: A visualization of the speech (black) and overlap-masking (gray) activity in a sentence stimulus in a simulated reverberant condition (left) and recorded a reverberant condition (right). Simulated RIRs assume that a surface's absorption coefficients are both frequency- and angle- independent, while the absorption coefficients of real-world RIRs depend on both frequency and angle. Therefore, the sentence in a simulated RIR (left) contains comparable quantities of overlap-making throughout the frequency channels, while the overlap-masking in the sentence in a recorded RIR (right) decreases with increasing frequency.

A quantification of the overlap-masking present in each channel is shown in Figure 6.10. This figure plots the electrode-specific percentages of pulses that were labeled as overlap-masking for CUNY sentences in a simulated RIR (blue) as well as in a recorded RIR (green). Both sentences recorded by male and female speakers were considered [73]. This figure demonstrates that for higher numbered electrodes (lower frequencies), similar amounts of overlap-masking are present in simulated and recorded reverberant conditions. Conversely, for the lower numbered electrodes, sentences exposed to simulated RIRs tend to contain more overlap-masking pulses than those exposed to recorded RIRs.

It should also be noted that, although a general decay in the percentage of overlap-masking pulses is evident in the higher numbered electrodes in both simulated and

recorded RIRs, this does not result from decreased reverberation in these channels. Rather, as shown in Figure 6.9, higher numbered electrodes often contain more speech stimuli than lower numbered electrodes, resulting in overlap-masking segments that are interrupted by a following speech token before they decay completely. Such interruptions are most likely responsible for the decrease in overlap-masking pulses in the higher numbered electrode channels.

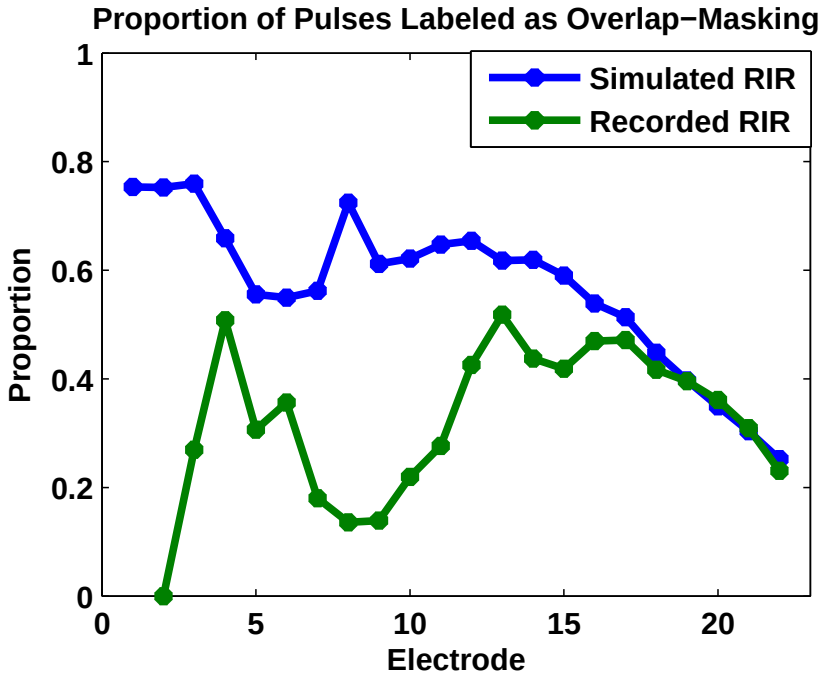

FIGURE 6.10: Quantification of the amount of overlap-masking pulses present in each channel in simulated reverberant conditions (blue) and recorded reverberant conditions (green). Similar amounts of overlap-masking are present in simulated and recorded conditions in the higher numbered electrodes (lower frequencies). Conversely, sentences exposed to simulated RIRs tend to contain more overlap-masking in the lower numbered electrodes than those exposed to recorded RIRs. The discrepancy results from the simulated RIR assumption that each surface's absorption coefficients are both frequency- and angle- independent, while real-world absorption coefficients depend on both frequency and angle.

The discrepancies can be explained by the methods used to generate simulated RIRs. In real world reverberant environments, each surface's reflection coefficients

are dependent on both frequency and angle. However, the simulated RIRs were generated assuming that the reflection coefficients are both frequency- and angle- independent [55]. Because real-world absorption is greater for higher frequencies (low electrode numbers), the simulated RIRs result in an unnatural amount of reverberation in the high frequency regions.

Although algorithm operating points were selected for each electrode in the experiments outlined in both Sections 6.1 and 6.2, the P_Ds that determined these operating points were constant across electrodes. One hypothesis is that reverberation mitigation may successfully improve speech recognition in recorded reverberant environments if P_Ds were allowed to differ between electrodes. Specifically, higher frequency bands may benefit from a less aggressive reverberation mitigation threshold. This is hypothesized because not only do these channels contain less overlap masking than the lower frequency channels, but they also often contain consonants which tend to be masked by higher energy vowels.

Because a follow-up experiment may require systematically varying the algorithm thresholds for all 22 electrodes, a large sentence database must first be developed. Once a new database has been developed, a thorough experiment can be conducted, aimed at selecting subject- and electrode- specific thresholds that result in improved speech recognition performance in mitigated reverberation.

The NH experiments presented in Sections 6.1 and 6.2 were conducted because NH listeners are more readily available than CI users and because the similarity in their speech recognition allows results to be pooled across listeners. Although NH studies can provide insight into an algorithm's effect on CI speech recognition, a study involving cochlear implant listeners is required to understand the full impact. Therefore, Section 6.4 investigates CI speech recognition in simulated RIRs with and without reverberation mitigation.

6.4 Cochlear Implant Sentence Recognition in Simulated RIRs

Reverberation mitigation, as presented in Section 6.1, successfully improved speech recognition in simulated reverberant conditions for normal hearing listeners using an acoustic model. To investigate the mitigation strategy's efficacy on mitigating reverberation for CI listeners using their own devices, a subsequent study was conducted.

6.4.1 Methods

Four CI subjects were recruited to study the effect of reverberation mitigation on speech recognition in simulated RIRs. Initially, speech was presented with simulated reverberation conditions with RTs set to 0.5s, 1.0s, and 1.5s, in a room with a dimension set to (10.0 x 6.6 x 3.0)m, a source located at (2.4835 x 2.0 x 1.8)m, and a microphone positioned at (6.5 x 3.8 x 1.8)m. However, after conducting three repeats with one subject, it was determined that a reverberation time of 1.5s was too high to adequately test speech recognition, as floor effects resulted. Subsequently, reverberation with RTs of 0.5s, 0.7s, and 1.0s were presented to the listeners.

Because each CI subject's results cannot be pooled due to differences in etiology and the relative physiological health of the neural population, five repeats per condition, with one sentence list presented per repeat, were conducted. Therefore, only three P_Ds were used to compare speech recognition between unmitigated reverberant speech and speech after reverberation mitigation. The P_D values of 30%, 45%, and 60%, were presented in random order. This experiment used the professionally recorded CUNY sentence lists, and reverberation was the only noise added to the tokens. Speech in SSN and reverberation and multi-talker babble and reverberation could not be presented, as the CUNY database does not consist of enough sentences lists.

6.4.2 Results

The results of the CI speech intelligibility study using simulated RIRs are displayed in Figure 6.11, plotted as percent of correctly identified speech. Results from each subject are displayed in separate subplots, and each group of data represents the presentation of a different reverberation time. Within the groups, performance in unmitigated reverberation is displayed in blue, while speech intelligibility after the application of reverberation mitigation is displayed at varying P_Ds in the remaining bars. Statistical significance is indicated by a line connecting two bars.

Subject S1 conducted five repeats at reverberation times of 0.5s and 1.0s. Due to the low speech intelligibility at an RT of 1.5s, only three repeats were conducted in this condition. Subsequently, two repeats were conducted using an RT of 0.7s. For the remaining subjects, five repeats were presented per condition. Figure 6.11 suggests that speech recognition performance trends are not consistent across subjects and algorithm thresholds (P_Ds).

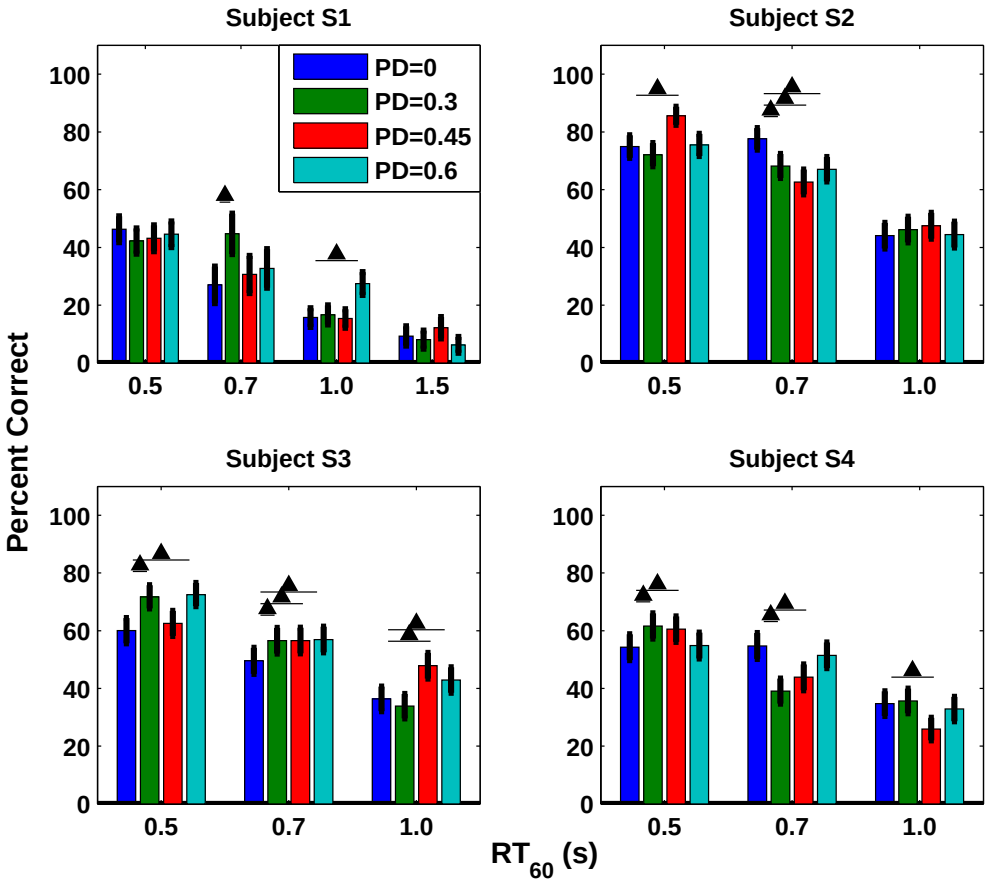

FIGURE 6.11: Speech recognition performance for four CI subjects (in separate subplots) in 3-4 reverberation conditions in both unmitigated reverberation (blue) and speech after reverberation mitigation (green, red, cyan). Speech recognition did not consistently improve across subjects and P_Ds after the application of reverberation mitigation.

Unlike the normal hearing listeners, whose similarity in speech recognition performance allows the results to be pooled across subjects, performance of CI listeners is more subject-dependent, requiring a separate study to be conducted for each subject. Many subject-specific differences, including duration of deafness, nerve survival, electrode placement, and bone growth, may affect the way sound is perceived. Therefore, a "one size fits all" approach to algorithm threshold selection may be less applicable to the CI population.

Specifically, CI listeners may benefit from P_D values that vary between electrodes. For example, depending on the perception resulting from each channel, one electrode may benefit from a more aggressive P_D of 60%, while another electrode may result in stronger performance using a P_D of 30%. Additional experiments would be required to test subject- and electrode- specific thresholds in CI listeners. To conduct this testing, a more complete sentence database would be required.

6.5 Discussion

The experiments in this chapter studied the efficacy of the reverberation mitigation algorithm in multiple reverberant environments for both NH listeners with stimuli presented with an acoustic model and for CI listeners using their own devices. In simulated reverberant conditions, speech recognition for NH listeners improved after the application of reverberation mitigation for several P_Ds. As P_D initially increased from a value of 0 (unmitigated reverberation), speech recognition tended to increase. However, once the P_D passed a critical point, speech recognition began to decrease, likely due to too many speech stimuli being removed. Based on these results, P_Ds of 30%, 45%, and 60% were selected for subsequent studies.

Three additional NH studies were conducted to examine the effects of reverberation mitigation in recorded reverberant environments, as well as in the presence of reverberation and additional noise. The mitigation strategy did not consistently improve speech recognition across subjects. However, a comparison of the effects of recorded and simulated RIRs on sentence stimuli highlighted uneven amounts of reverberation in different frequency channels in the presence of recorded RIRs, compared to more consistent frequency-specific reverberation levels in simulated reverberation. This discrepancy suggests that channel-specific P_Ds may improve the efficacy of reverberation mitigation in recorded RIRs. Future work is required to test this hypothesis.

Next, a study was conducted that presented CI listeners with speech in simulated reverberant conditions, both before and after reverberation mitigation. Speech recognition did not consistently improve for the four individual subjects. Unlike NH subjects, whose speech recognition similarities allow for the pooling of data, many subject-specific variables effect CI speech recognition. A few of these variables include duration of deafness, nerve survival, electrode array insertion depth, and bone growth. Such subject- and channel- specific properties suggest that CI listeners may benefit from a reverberation mitigation strategy that is tailored to both subject and electrode. Future work requiring a more extensive sentence database is necessary to test this hypothesis.

7

Conclusions

This research project aimed to detect and mitigate reverberation effects in cochlear implant pulse trains. Initially, it was demonstrated that reverberant speech can be successfully discriminated from speech in quiet, SSN, or WGN. Reverberation detection appeared to be robust to both subject clinical parameters as well as to most environment conditions. Reverberation detection was an important first step towards reverberation mitigation, as a detection algorithm can be used to initiate the mitigation strategy. This initiation allows the mitigation strategy to be tailored to reverberant speech, minimizing its effect on quiet stimuli.

It was hypothesized that the design of a mitigation algorithm that focused on reverberation effects such as self-masking or overlap-masking would allow for the design of a causal mitigation algorithm. While other mitigation algorithms proposed in the literature have shown potential benefits, these algorithms rely on non-causal features in an attempt to estimate the quiet signal. In order to determine if the hypothesized approach might be feasible, an initial study was conducted using both NH and CI listeners, investigating the impact of ideally mitigating self- and overlap-masking effects independently. This study found that mitigating either effect had

the potential to benefit speech recognition in reverberant environments, suggesting that the hypothesized approach was feasible.

Although mitigating either self- or overlap- masking resulted in improved speech recognition, mitigating self-masking would require correcting for amplitude corruption during active speech, which may be difficult for real-time implementation. Overlap-masking mitigation, on the other hand, does not occur during active speech segments and would require detecting and removing the masking pulses. Therefore, this study developed two overlap-masking mitigation strategies, one operating on the acoustic signal, and the other operating on the CI-processed signal. Because performance was comparable between the two methods, only the CI-based mitigation strategy was implemented into a speech processing algorithm.

To test the efficacy of the reverberation mitigation strategy, an initial study was conducted with NH subjects and an acoustic model. Listeners were presented with speech in simulated reverberant conditions both with- and without- reverberation mitigation, and their speech recognition was studied. Speech intelligibility was found to improve after reverberation mitigation, and P_Ds of 30%, 45%, and 60% were selected as the operating points for subsequent experiments.

Next, three studies presented NH listeners with reverberant speech using recorded RIRs, and speech recognition in mitigated reverberation was compared to that in unmitigated reverberation. The initial study presented CUNY sentences recorded by amateur speakers, while the following two studies used professional recordings. Additionally, different methods of threshold selection were investigated: one that assumed no prior knowledge of the RIRs, and another with access to the reverberation parameters. However, in all three studies, reverberation mitigation did not consistently improve speech recognition performance.

One hypothesis to explain why speech recognition performance did not improve in recorded reverberant conditions considers the effects of both simulated and recorded

RIRs on speech stimuli. In recorded reverberant environments, reverberation is dependent on both frequency and angle, with higher frequencies experiencing less reverberation. In simulated environments, on the other hand, reverberation effects are assumed to be both frequency- and angle- independent, resulting in similar levels of overlap-masking in all frequency channels. Because higher frequency channels in recorded RIRs experience less reverberation, and therefore less overlap-masking effects, the mitigation algorithm may benefit from electrode-specific P_Ds.

Future work is required to test whether electrode-specific P_Ds could improve speech recognition in mitigated reverberation in recorded RIRs. Recall that P_Ds of 30%, 45%, and 60% resulted in speech recognition improvements in simulated RIRs, as presented in Section 6.1. Also recall that similar quantities of overlap-masking are present in high numbered electrodes in both simulated and recorded RIRs (see Figure 6.10). Therefore, it is hypothesized that a P_D of 30%, 45%, or 60% should be selected for electrodes 19-22. Because less overlap-masking is present in the lower numbered electrodes, it is hypothesized that P_Ds of equal or lesser values should be selected for electrodes 1-18. An initial study to test this hypothesis could present NH listeners with unmitigated and mitigated reverberation using electrode-specific P_Ds. To reduce experimental conditions, P_Ds could be assigned in groups of 3-4 electrodes, such that electrodes 1-3, 4-6, 7-10, 11-14, 15-18, and 19-22 would be assigned the same value. P_Ds for electrodes 1-18 could vary between 15% and 60%, and P_Ds for electrodes 19-22 could vary between 30% and 60%. Speech recognition using various combinations of electrode-specific P_Ds could be collected, with the goal of improving speech intelligibility in mitigated reverberant conditions using recorded RIRs.

As a final experiment in this work, a CI speech intelligibility study was conducted to determine whether the mitigation algorithms can improve speech recognition for CI listeners in reverberant environments. Because speech recognition did not consistently improve for CI listeners after reverberation mitigation, future work requires

determining whether subject- and electrode- specific mitigation parameters may impact performance.

Unlike NH listeners presented with an acoustic model, many subject-specific variables affect CI listeners' speech recognition. Further, the perception resulting from each electrode in a CI listener's electrode array can vary based on nerve survival and bone growth, as well as other factors. For this reason, a future study should investigate the effect of selecting subject- and electrode- specific P_Ds for reverberation mitigation in CI listeners. Similarly to the proposed NH study, electrode-specific P_Ds could be assigned in groups of 3-4 electrodes. Assuming an initial study is conducted with CI listeners and simulated RIRs, P_Ds for each electrode-group could vary from 30% to 60%. If electrode-specific P_Ds are found that improve speech recognition, a follow-up study should be conducted presenting CI listeners with mitigated and unmitigated reverberation in recorded RIRs.

Bibliography

[1] K. Kokkinakis and P. Loizou, "The impact of reverberant self-masking and overlap-masking effects on speech intelligibility by cochlear implant listeners (L)," *J. Acoust. Soc. Am.*, vol. 130(3), pp. 1099–1102, 2011.

[2] R. Bolt and A. MacDonald, "Theory of speech masking by reverberation," *J. Acoust. Soc. Am.*, vol. 21(6), pp. 577–580, 1949.

[3] A. Nabelek and T. Letowski, "Similarities of vowels in nonreverberant and reverberant fields," *J. Acoust. Soc. Am.*, vol. 83(5), pp. 1891–1899, 1988.

[4] A. Nabelek, T. Letowski, and F. Tucker, "Reverberant overlap- and self-masking in consonant identification," *J. Acoust. Soc. Am.*, vol. 86(4), pp. 1259–1265, 1989.

[5] A. Kjellberg, "Effects of reverberation time on the cognitive load in speech communication: Theoretical considerations," *Noise Health*, vol. 7(25), pp. 11–21, 2004.

[6] T. Finitzo-Hieber and T. Tillman, "Room acoustics effects on monosyllabic word discrimination ability for normal and hearing-impaired children," *J. Speech Hear. Res.*, vol. 21(3), pp. 440–458, 1978.

[7] K. Kokkinakis, O. Hazrati, and P. Loizou, "A channel-selection criterion for suppressing reverberation in cochlear implants," *J. Acoust. Soc. Am.*, vol. 129(5), pp. 3221–3232, 2011.

[8] M. Skinner, L. Holden, L. Whitford, K. Plant, C. Psarros, and T. Holden, "Speech recognition with the Nucleus 24 SPEAK, ACE, and, CIS speech coding strategies in newly implanted adults," *Ear Hear.*, vol. 23(3), pp. 207–223, 2002.

[9] A. Spahr and M. Dorman, "Performance of subjects fit with the Advanced Bionics CII and Nucleus 3G cochlear implant devices," *Arch. Otolaryngol. Head Neck Surg.*, vol. 130(5), pp. 624–628, 2004.

[10] J. Allen and D. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65(4), pp. 943–950, 1979.

[11] S. Neely and J. Allen, "Invertibility of a room impulse response," *J. Acoust. Soc. Am.*, vol. 66(1), pp. 165–169, 1979.

[12] J. Mourjopoulos, P. Clarkson, and J. Hammond, "A comparitive study of least-squares and homomorphic techniques for the inversion of mixed phase signals," *Proc. ICASSP*, vol. 7, pp. 1858–1861, 1982.

[13] M. Tohyama, R. Lyon, and T. Koike, "Source waveform recovery in a reverberant space by cepstrum dereverberation," *Proc. ICASSP*, vol. 1, pp. 157–160, 1993.

[14] J. Mourjopoulos, "Digital equalization of room acoustics," *J. Audio Eng. Soc.*, vol. 42(11), pp. 884–900, 1994.

[15] R. Kennedy and B. Radlović, "Iterative cepstrum-based approach for speech dereverberation," *Proc. ISSPA*, vol. 1, pp. 55–58, 1999.

[16] B. Radlović and R. Kennedy, "Nonminimum-phase equalization and its subjective importance in room acoustics," *IEEE Trans. Speech Audio Process.*, vol. 8(6), pp. 728–737, 2000.

[17] K. Kokkinakis and P. Loizou, "Selective-tap blind dereverberation for two-microphone enhancement of reverberant speech," *IEEE Signal Process. Lett.*, vol. 16, pp. 961–964, 2009.

[18] O. Hazrati and P. Loizou, "Reverberation suppression in cochlear implants using a blind channel-selection strategy," *J. Acoust. Soc. Am.*, vol. 133(6), pp. 4188–4196, 2013.

[19] O. Hazrati, J. Lee, and P. Loizou, "Blind binary masking for reverberation suppression in cochlear implants," *J. Acoust. Soc. Am.*, vol. 133(3), pp. 1607–1614, 2013.

[20] J. Desmond, L. Collins, and C. Throckmorton, "Using channel-specific statistical models to detect reverberation in cochlear implant stimuli," *J. Acoust. Soc. Am.*, vol. 134(2), pp. 1112–1120, 2013.

[21] J. Desmond, L. Collins, and C. Throckmorton, "Determining the effects of reverberant self- and overlap- masking on speech recognition for cochlear implant listeners," *JASA-EL*, vol. 135(6), pp. 304–310, 2014.

[22] H. Schuknecht, *Pathology of the Ear*. Lea and Febiger, 1993.

[23] P. Loizou, "Introduction to cochlear implants," *IEEE Signal Process. Mag.*, vol. 18(1), pp. 32–42, 1999.

[24] F. Zeng, S. Rebscher, W. Harrison, X. Sun, and H. Feng, "Cochlear implants: system design, integration, and evaluation," *IEEE Rev. Biomed. Eng.*, vol. 1, pp. 115–142, 2008.

[25] H. Kuttruff, *Room Acoustics*. Spon Press, 2009.

[26] W. Sabine, *Collected Papers on Acoustics*. Harvard University Press, 1922.

[27] T. Holman, *Sound for Film and Television*. Focal Press, 2010.

[28] A. Nabelek and P. Dagenais, "Vowel errors in noise and in reverberation by hearing-impaired listeners," *J. Acoust. Soc. Am.*, vol. 80(3), pp. 741–748, 1986.

[29] A. Nabelek and P. Robinson, "Monaural and binaural speech perception in reverberation for listeners of various ages," *J. Acoust. Soc. Am.*, vol. 71(5), pp. 1242–1248, 1982.

[30] L. Lamel, R. Kassel, and S. Seneff, "Speech database development: Design and analysis of the acoustic-phonetic corpus," *Proc. DARPA Speech Recog. Workshop*, pp. 100–109, 1986.

[31] K. Furuya and A. Kakaoka, "Robust speech dereverberation using multichannel blind deconvolution with spectral subtraction," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, pp. 1579–1591, 2007.

[32] P. Krishnamoorthy and S. Prasanna, "Reverberant speech enhancement by temporal and spectral processing," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, pp. 253–266, 2009.

[33] M. Miyoshi and Y. Kaneda, "Inverse filtering of room acoustics," *IEEE Trans. Speech Audio Process.*, vol. 36(2), pp. 145–152, 1988.

[34] Y. Huang, J. Benesty, and J. Chen, *Springer Handbook of Speech Processing*. Springer, New York, 2007.

[35] Y. Lin and D. Lee, "Bayesian regularization and nonnegative deconvoluation for room impulse response estimation," *Noise Health*, vol. 7(25), pp. 11–21, 2006.

[36] P. Naylor and N. Gaubitch, eds., *Speech Dereverberation*. Springer London, 2010.

[37] M. Schroeder, "Integrated-impulse method for measuring sound decay without using impulses," *J. Acoust. Soc. Am.*, vol. 66, pp. 497–500, 1979.

[38] C. Dunn and M. Hawksford, "Distortion of immunity of MLS-derived impulse response measurements," *J. Acoust. Eng. Soc.*, vol. 41, pp. 314–335, 1993.

[39] N. Ream, "Nonlinear identification using inverse-repeat m sequences," *Proc. IEEE*, vol. 117(1), pp. 213–218, 1970.

[40] P. Briggs and K. Godfrey, "Pseudorandom signals for the dynamic analysis of multivariable systems," *Proc. IEEE*, vol. 113, pp. 1259–1267, 1966.

[41] N. Aoshima, "Computer-generated pulse signal applied for sound measurement," *J. Acoust. Soc. Am.*, vol. 69(5), pp. 1484–1488, 1981.

[42] Y. Suzuki, F. Asano, H.-Y. Kim, and T. Sone, "An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses," *J. Acoust. Soc. Am.*, vol. 97(2), pp. 1119–1123, 1995.

[43] A. Berkhout, M. Boone, and C. Kesselman, "Acoustic impulse response measurement: A new technique," *J. Aud. Eng. Soc.*, vol. 32(10), pp. 740–746, 1984.

[44] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," *In: 108th AES Convention, Paris (France)*, 2000.

[45] A. Keshavarz, S. Mosayyebpour, M. Biguesh, T. Gulliver, and M. Esmaeili, "Speech-model based accurate blind reverberation time estimation using an LPC filter," *IEEE Trans. Audio Speech Lang. Process.*, vol. 206, pp. 1884–1893, 2012.

[46] M. Schroeder, "New method of measuring reverberation time," *J. Acoust. Soc. Am.*, vol. 66, pp. 1187–1188, 1965.

[47] M. Wu and D. Wang, "A pitch-based method for the estimation of short reverberation time," *Acta Acustica*, vol. 92, pp. 337–339, 2006.

[48] M. Unoki and S. Hiramatsu, "Blind estimation method of reverberation time based on concept of modulation transfer function," *J. Acoust. Soc. Am.*, vol. 123(5), pp. 3616–3616, 2008.

[49] J. Wen, E. Habets, and P. Naylor, "Blind estimation of reerberation time based on the distribution of signal decay rates," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process*, pp. 329–332, March-April 2008.

[50] R. Ratnam, D. Jones, B. Wheeler, W. O'Brien Jr., C. Lansing, and A. Feng, "Blind estimation of reverberation time," *J. Acoust. Soc. Am.*, vol. 114(5), pp. 2877–2892, 2003.

[51] B. Yegnarayana and P. Murthy, "Enhancement of reverberant speech using LP residual signal," *IEEE Trans. Speech Audio Process.*, vol. 8(3), pp. 267–281, 2000.

[52] N. Gaubitch, D. Ward, and P. Naylor, "Statistical analysis of the autoregresive modeling of reverberant speech," *J. Acoust. Soc. Am.*, vol. 120(6), pp. 4031–4039, 2006.

[53] B. Gillespie, H. Malvar, and D. Florencio, "Speech dereverberation via maximum-kurtosis subband adaptive filtering," *ICASSP*, pp. 3701–3704, 2001.

[54] M. Wu and D. Wang, "A two-stage algorithm for one-microphone reverberant speech enhancement," *IEEE Trans. Audio Speech Lang. Process.*, vol. 14(3), pp. 774–784, 2006.

[55] E. Lehmann and A. Johansson, "Prediction of energy decay in room impulse responses simulated with an image-source model," *J. Acoust. Soc. Am.*, vol. 124(1), pp. 269–277, 2008.

[56] M. Nilsson, S. D. Soli, and J. A. Sullivan, "Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise," *J. Acoust. Soc. Am.*, vol. 95(2), pp. 1085–1099, 1994.

[57] C. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[58] M. Tipping, "Sparse bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, 2001.

[59] B. Champagne, S. Bedard, and A. Stephenne, "Performance of time-delay estimation in the presence of room reverberation," *IEEE Trans. Speech Audio Process.*, vol. 4(2), pp. 148–152, 1996.

[60] P. Blamey, R. Dowell, Y. Tong, and G. Clark, "An acoustic model of a multiple-channel cochlear implant," *J. Acoust. Soc. Am.*, vol. 76(1), pp. 97–103, 1984.

[61] H. Davis and S. Silverman, *Hearing and Deafness, ed 4 (Central Institute for the Deaf sentence lists)*. Holt, Rhinehart, & Winston, 1978.

[62] S. Tantum, L. Collins, and C. Throckmorton, "Bayesian a posteriori performance estimation for speech recognition and psychophysical tasks," *Submitted to: J. Acoust. Soc. Am.*, 2013.

[63] D. Kewley-Port, T. Burkle, and J. Lee, "Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners," *J. Acoust. Soc. Am.*, vol. 122(4), pp. 2365–2375, 2007.

[64] M. Breeuwer and R. Plomp, "Speechreading supplemented with frequency-selective sound-pressure information," *J. Acoust. Soc. Am.*, vol. 76(3), pp. 686–691, 1984.

[65] M. Breeuwer and R. Plomp, "Speechreading supplemented with formant frequency information from voiced speech," *J. Acoust. Soc. Am.*, vol. 77(1), pp. 314–317, 1985.

[66] M. Breeuwer and R. Plomp, "Speechreading supplemented with auditorily-presented speech parameters," *J. Acoust. Soc. Am.*, vol. 79(2), pp. 481–499, 1986.

[67] R. Lippmann, "Accurate consonant perception without mid-frequency energy," *IEEE Trans. Speech Audio. Proc.*, vol. 4(1), pp. 66–69, 1996.

[68] R. Warren, K. Reiner, J. Bashford Jr., and B. Braubaker, "Spectral redundancy: Intelligibility of sentences heard through narrow spectral slits," *Percept. Psychophys.*, vol. 57(2), pp. 175–182, 1995.

[69] R. Shannon, J. Galvin III, and D. Baskent, "Holes in hearing," *J. Assoc. Res. Oto.*, vol. 3, pp. 185–199, 2001.

[70] T. Houtgast and H. Steeneken, "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria," *J. Acoust. Soc. Am.*, vol. 77(3), pp. 1069–1077, 1985.

[71] Y. Hu and P. Loizou, "A new sound coding strategy for suppressing noise in cochlear implants," *J. Acoust. Soc. Am.*, vol. 124(1), pp. 498–509, 2008.

[72] M. Jeub, M. Schäfer, and P. Vary, "A binaural room impulse response database for the evaluation of dereverberation algorithms," *Proc. 16th Int. Conf. Digital Signal Process.*, 2009, Santorini, Greece.

[73] A. Boothroyd, L. Hanin, and T. Hnath, *A sentence test of speech perception: Reliaility, set equivalence, and short term lerning.* Internal Report RCI 10, New York: City University of New York, 1985.

Biography

Jill Desmond was born on June 7, 1986 and grew up in the Boston suburb of Marshfield, MA. She graduated salutatorian from Marshfield High School in 2005. In 2009, Jill graduated *cum laude* from the University of Delaware with an honors B.E.E. degree and a mathematics minor. She went on to receive her M.S. in electrical and computer engineering from Duke University in 2011, completing a thesis entitled "Using Forward Masking Patterns to Predict Imperceptible Information in Speech for Cochlear Implant Subjects." Jill received her Ph.D. from Duke University in 2014, defending the dissertation "Using Channel-Specific Models to Detect and Mitigate Reverberation in Cochlear Implants." Both her M.S. and Ph.D. degrees were completed under the supervision of Dr. Leslie Collins and Dr. Chandra Throckmorton.